



CONCISE ENCYCLOPEDIA OF LANGUAGES OF THE WORLD

KEITH BROWN • SARAH OGILVIE



CONCISE ENCYCLOPEDIA OF LANGUAGES OF THE WORLD

This page intentionally left blank

Presented by: <https://jafrilibrary.com>

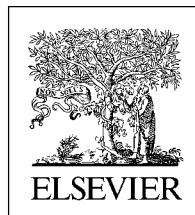
CONCISE ENCYCLOPEDIA OF LANGUAGES OF THE WORLD

COORDINATING EDITOR

KEITH BROWN
University of Cambridge
Cambridge, UK

CO-EDITOR

SARAH OGILVIE
University of Oxford
Oxford, UK



Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

Elsevier Ltd., The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

© 2009 Elsevier Ltd.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic, or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publishers.

Permissions may be sought directly from Elsevier's Rights Department in Oxford, UK: phone (+44) 1865 843830; fax (+44) 1865 853333; e-mail permissions@elsevier.com.

Requests may also be completed online via the homepage (<http://www.elsevier.com/locate/permissions>).

First edition 2009

Library of Congress Control Number: 2008934269

A catalogue record for this book is available from the British Library

ISBN 978-0-08-087774-7

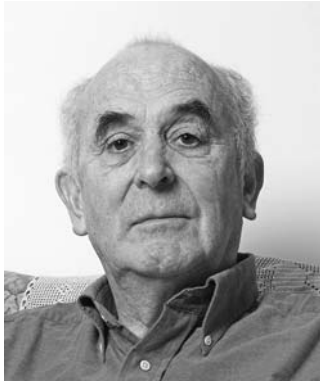
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

This book is printed on acid-free paper
Printed and bound in China

Cover image: Adapted from *Orbis Terrarum Nova* (1594) by Petro Plancio



THE EDITORS



Keith Brown was Editor-in-Chief of the second edition of the *Encyclopedia of Language and Linguistics* (Elsevier, 2006). He is now an Associate Lecturer in the Faculty of English at Cambridge. From 2007 he has been President of the Philological Society. From 1990 to 1994 he was President of the Linguistics Association of Great Britain, and he has been a Member of Council of the Philological Society since 1998. He is author of *Linguistics Today* (Fontana, 1984) and co-author, with Jim Miller, of *Syntax: A Linguistic Introduction to Sentence Structure* and *Syntax: Generative Grammar* (Hutchinson, 1981). Keith was joint editor of *Concise Encyclopedia of Linguistic Theories* and *Concise Encyclopedia of Grammatical Categories* (Pergamon Press, 1997 and 1998), *Common Denominators in Art and Science* (Aberdeen University Press, 1983) and *Language, Reasoning and Inference* (Academic Press, 1986).



Sarah Ogilvie, Trinity College, Oxford, is a linguist and lexicographer who specializes in words that enter English from non-European languages. She was Languages of the World section editor of the second edition of the *Encyclopedia of Language and Linguistics* (Elsevier, 2006), a former editor of the *Oxford English Dictionary*, and was Etymologies Editor of the *Shorter Oxford Dictionary* (6th ed., 2007).

This page intentionally left blank

ALPHABETICAL LIST OF ARTICLES

| | |
|-------------------------------------|--|
| Abkhaz | Balto-Slavic Languages |
| Adamawa-Ubangi | Bantu Languages |
| Africa as a Linguistic Area | Bashkir |
| Afrikaans | Basque |
| Afroasiatic Languages | Belorussian |
| Ainu | Bengali |
| Akan | Benue–Congo Languages |
| Akkadian | Berber |
| Albanian | Bikol |
| Algonquian and Ritwan Languages | Bislama |
| Altaic Languages | Brahui |
| Amharic | Breton |
| Anatolian Languages | Bulgarian |
| Ancient Egyptian and Coptic | Burmese |
| Andean Languages | Burushaski |
| Arabic | Caddoan Languages |
| Arabic as an Introflecting Language | Cape Verdean Creole |
| Arabic Languages, Variation in | Cariban Languages |
| Aramaic and Syriac | Catalan |
| Arawak Languages | Caucasian Languages |
| Areal Linguistics | Cebuano |
| Armenian | Celtic |
| Arrernte | Central Siberian Yupik as a Polysynthetic Language |
| Artificial Languages | Central Solomon Languages |
| Assamese | Chadic Languages |
| Australia: Language Situation | Chibchan |
| Australian Languages | Chimakuan Languages |
| Austic Hypothesis | Chinantec: Phonology |
| Austroasiatic Languages | Chinese |
| Austronesian Languages | Chinese as an Isolating Language |
| Austro-Tai Hypotheses | Choco Languages |
| Avestan | Chorasmian |
| Aymar  | Chukotko-Kamchatkan Languages |
| Azerbaijani | Church Slavonic |
| Bactrian | Chuvash |
| Balinese | Classification of Languages |
| Balkans as a Linguistic Area | Cornish |
| Balochi | Cree |

| | |
|---|----------------------------------|
| Creek | Greek, Ancient |
| Crow | Greek, Modern |
| Cupeño | Guaraní |
| Cushitic Languages | Gujarati |
| Czech | Gullah |
| Danish | Gur Languages |
| Dardic | Guugu Yimithirr |
| Dhivehi | Hausa |
| Diachronic Morphological Typology | Hawaiian |
| Dinka | Hawaiian Creole English |
| Dogon | Hebrew, Biblical and Jewish |
| Domari | Hebrew, Israeli |
| Dravidian Languages | Highland East Cushitic Languages |
| Dutch | Hiligaynon |
| Eblaite | Hindi |
| Efik | Hindustani |
| Elamite | Hiri Motu |
| Endangered Languages | Hittite |
| English in the Present Day | Hmong-Mien Languages |
| English, African-American Vernacular | Hokan Languages |
| English, Early Modern | Hopi |
| English, Later Modern (ca. 1700–1900) | Hungarian |
| English, Middle English | Hurrian |
| English, Old English | Ijò |
| English, Variation in Nonnative Varieties | Ilocano |
| English: World Englishes | Indo-Aryan Languages |
| Eskimo–Aleut | Indo-European Languages |
| Esperanto | Indo-Iranian |
| Estonian | Inupiaq |
| Ethiopia as a Linguistic Area | Iranian Languages |
| Ethiopian Semitic Languages | Iroquoian Languages |
| Ethnologue | Italian |
| Etruscan | Italian as a Fusional Language |
| Europe as a Linguistic Area | Italic Languages |
| Evenki | Japanese |
| Ewe | Javanese |
| Fanagalo | Jèrriais |
| Fijian | Jewish Languages |
| Finnish | Jiwarli |
| Finnish as an Agglutinating Language | Kalkutungu |
| Flores Languages | Kannada |
| Formosan Languages | Kanuri |
| Franglais | Kapampangan |
| French | Karen Languages |
| Fulfulde | Kashmiri |
| Galician | Kayardild |
| Gamilaraay | Kaytetye |
| Gə'əz | Kazakh |
| Georgian | Keres |
| German | Ket |
| Germanic Languages | Khasi |
| Gikuyu | Khmer |
| Goidelic Languages | Khoesaaan Languages |
| Gondi | Khotanese |
| Gothic | Kinyarwanda |

| | |
|---|-----------------------------------|
| Kirghiz | Ngan'gi |
| Kordofanian Languages | Niger-Congo Languages |
| Korean | Nilo-Saharan Languages |
| Krio | Niuean |
| Kru Languages | Nivkh |
| Kurdish | Norse and Icelandic |
| Kurukh | North Philippine Languages |
| Kwa Languages | Norwegian |
| Lahnda | Nostratic Hypothesis |
| Lak | Nuristani Languages |
| Lakota | Nuuchahnulth |
| Lao | Nyanja |
| Latin | Occitan |
| Latvian | Old Church Slavonic |
| Lithuanian | Omaha-Ponca |
| Long-Range Comparison: Methodological Disputes | Omotoc Languages |
| Louisiana Creole | Oneida |
| Luganda | Oromo |
| Luo | Ossetic |
| Luxembourgish | Oto-Mangean Languages |
| Macedonian | Pahlavi |
| Macro-Jê | Palenquero |
| Madang Languages | Pāli |
| Madurese | Panoan Languages |
| Malagasy | Papiamentu |
| Malay | Papuan Languages |
| Malayalam | Pashto |
| Malayo-Polynesian Languages | Persian, Modern |
| Maltese | Persian, Old |
| Malukan Languages | Phoenician |
| Mambila | Pictish |
| Manambu | Pidgins and Creoles |
| Mande Languages | Pidgins and Creoles, Variation in |
| Maori | Pitjantjatjara/Yankunytjatjara |
| Mapudungan | Polish |
| Marathi | Pomoan Languages |
| Mayan Languages | Portuguese |
| Michif | Punjabi |
| Misumalpan | Quechua |
| Mixe-Zoquean Languages | Rhaeto Romance |
| Mobilian Jargon | Riau Indonesian |
| Mon | Romance Languages |
| Mongolic Languages | Romani |
| Mon-Khmer Languages | Romanian |
| Morphological Typology | Russenorsk |
| Morrobalama | Russian |
| Munda Languages | Ryukyuan |
| Muskogean Languages | Saami |
| Na-Dene Languages | Salishan Languages |
| Nahuatl | Samar-Leyte |
| Native American Languages | Sango |
| Native Languages of North America, Variation in | Sanskrit |
| Navajo | Santali |
| Nenets | Scots |
| Nepali | Scots Gaelic |

x **Alphabetical List of Articles**

| | |
|---|--|
| Semitic Languages | Tocharian |
| Serbian-Croatian-Bosnian Linguistic Complex | Toda |
| Shona | Tohono O'odham |
| Sign Languages | Tok Pisin |
| Sign Language: Morphology | Torricelli Languages |
| Sign Languages of the World | Totonacan Languages |
| Sindhi | Trans New Guinea Languages |
| Sinhala | Tsotsi Taal |
| Sino-Tibetan Languages | Tucanoan Languages |
| Siouan Languages | Tungusic Languages |
| Skou Languages | Tupian Languages |
| Slavic Languages | Turkic Languages |
| Slovak | Turkish |
| Slovene | Turkmen |
| Sogdian | Ugaritic |
| Somali | Ukrainian |
| Songhay Languages | United States of America: Language Situation |
| Sorbian | Uralic Languages |
| South Asia as a Linguistic Area | Urdu |
| South Philippine Languages | Uto-Aztecan Languages |
| Southeast Asia as a Linguistic Area | Uyghur |
| Southern Bantu Languages | Uzbek |
| Spanish | Vietnamese |
| Sumerian | Vurës |
| Swahili | Wa |
| Swedish | Wakashan |
| Syriac | Wambaya |
| Tagalog | Warlpiri |
| Tahitian | Welsh |
| Tai Languages | West Greenlandic |
| Tajik Persian | West Papuan Languages |
| Tamambo | Wolaitta |
| Tamil | Wolof |
| Tanoan | Xhosa |
| Tariana | Yakut |
| Tatar | Yanito |
| Telugu | Yiddish |
| Thai | Yoruba |
| Tibetan | Yukaghir |
| Tigrinya | Zapotecan |
| Tiwi | Zulu |

SUBJECT CLASSIFICATION

Note that italicized titles are included for classification purposes only and do not cross-refer to articles.

Introduction

List of Abbreviations

Classification of Languages

Areal Linguistics

Africa as a Linguistic Area

Balkans as a Linguistic Area

Ethiopia as a Linguistic Area

Europe as a Linguistic Area

South Asia as a Linguistic Area

Southeast Asia as a Linguistic Area

Afroasiatic Languages

Ancient Egyptian and Coptic

Berber Languages

Chadic Languages

Hausa

Cushitic Languages

Highland East Cushitic Languages

Oromo

Somali

Omotic Languages

Wolaitta

Semitic Languages

Eblaite

Central

Arabic

Arabic Languages, Variation in

Aramaic and Syriac

Hebrew, Biblical and Jewish

Hebrew, Israeli

Jewish languages

Maltese

Phoenician

Syriac

Ugaritic

Eastern

Akkadian

Southern

Ethiopian Semitic Languages

Amharic

Gə'əz

Tigrinya

Altaic Languages

Mongolic Languages

Tungusic Languages

Evenki

Turkic Languages

Azerbaijani

Bashkir

Chuvash

Kazakh

Kirghiz

Tatar

Turkish

Turkmen

Uyghur

Uzbek

Yakut

Australian Languages

Australia: Language Situation

Mirndi

Wambaya

Pama-Nyungan

Arrernte

Gamilaraay

Guugu Yimithirr

Jiwarli

Kalkutungu

Kaytetye
Morrobalama
Pitjantjatjara / Yankunytjatjara
Warlpiri

Southern Daly

Ngan'gi
Tangkic
Kayardild
Tiwi

Austroasiatic Languages

Mon-Khmer Languages

Eastern

Khmer

Northern

Khasi

Vietnamese

Wa

Southern

Mon

Munda Languages

Santali

Austronesian Languages

Formosan Languages

Malayo-Polynesian Languages

Central

Flores Languages

Malukan Languages

Oceanic

Fijian

Maori

Tahitian

Tamambo

Vures

Western

Balinese

Bikol

Cebuano

Hawaiian

Hiligaynon

Ilocano

Javanese

Kapampangan

Madurese

Malagasy

Malay

Niuean

North Philippine Languages

Riau Indonesian

Samar-Leyte

South Philippine Languages

Tagalog

Caucasian Languages

Abkhaz

Georgian

Lak

Chukotko-Kamchatkan Languages

Dravidian Languages

Brahui

Gondi

Kannada

Kurukh

Malayalam

Tamil

Telugu

Toda

Hmong-Mien Languages

Indo-European Languages

Albanian

Anatolian Languages

Hittite

Armenian

Balto-Slavic Languages

Baltic Languages

Latvian

Lithuanian

Slavic Languages

Belorussian

Bulgarian

Church Slavonic

Czech

Macedonian

Old Church Slavonic

Polish

Russian

'Serbian-Croatian-Bosnian Linguistic Complex'

Slovak

Slovene

Sorbian

Ukrainian

Celtic

Breton

Cornish

Goidelic Languages

Pictish

Scots Gaelic

Welsh

Germanic Languages

Afrikaans

Danish

Dutch

English, Old English

English, Middle English

- English, Early Modern
 English, Later Modern (ca. 1700–1900)
 English in the Present Day
 English, World Englishes
 English, African-American Vernacular
 English, Variation in Nonnative Varieties
 German
 Gothic
 Luxembourgish
 Norse and Icelandic
 Norwegian
 Scots
 Swedish
 Yiddish
Hellenic
 Greek, Ancient
 Greek, Modern
Indo-Iranian
 Indo-Aryan Languages
 Assamese
 Bengali
 Dardic
 Kashmiri
 Dhivehi
 Domari
 Gujarati
 Hindi
 Hindustani
 Lahnda
 Marathi
 Nepali
 Nuristani Languages
 Pāli
 Punjabi
 Romani
 Sanskrit
 Sindhi
 Sinhala
 Urdu
 Iranian Languages
 Avestan
 Bactrian
 Balochi
 Chorasmian
 Khotanese
 Kurdish
 Ossetic
 Pahlavi
 Pashto
 Persian, Modern
 Persian, Old
 Sogdian
 Tajik Persian
Italic Languages
 Latin
- Romance Languages
 Catalan
 French
 Galician
 Italian
 Jèrriais
 Occitan
 Portuguese
 Rhaeto Romance
 Romanian
 Spanish
 Tocharian

Khoesaaan Languages
Niger-Congo Languages
 Adamawa-Ubangi
Atlantic Congo Languages
 Fulfulde
 Ijò
 Wolof
Benue-Congo Languages
 Efik
 Mambila
Bantu Languages
 Gikuyu
 Kinyarwanda
 Luganda
 Nyanja
 Shona
 Southern Bantu Languages
 Swahili
 Xhosa
 Zulu
 Dogon
 Gur Languages
 Kru Languages
 Kwa Languages
 Akan
 Ewe
 Yoruba
Kordofanian Languages
Mande Languages

Nilo-Saharan Languages
 Dinka
 Kanuri
 Luo
 Songhay Languages

North American and Middle American Languages
 Algonquian and Ritwan Languages
 Cree
 Michif
 Caddoan Languages
 Chibchan

Chimakuan Languages
Eskimo-Aleut
 Inupiaq
 West Greenlandic
Hokan Languages
 Pomoan Languages
Iroquoian Languages
 Oneida
Keres
Mayan Languages
Misumalpan
Mixe-Zoquean Languages
Muskogean Languages
 Creek
Na-Dene Languages
 Navajo
Native American Languages
 Native Languages of North America, Variation in
Oto-Mangean Languages
 Chinantec: Phonology
 Zapotecan
Salishan Languages
Siouan Languages
 Crow
 Lakota
 Omaha-Ponca
Tanoan Languages
Totonacan Languages
United States of America: Language Situation
Uto-Aztecan Languages
 Cupeño
 Hopi
 Nahuatl
 Tohono O'odham
Wakashan
 Nuuchahnulth

Papuan Languages

Central Solomon Languages
Manambu
Skou Languages
Torricelli Languages
Trans New Guinea Languages
 Madang Languages
West Papuan Languages

Pidgins and Creoles

Bislama
Cape Verdean Creole
Fanagalo
Franglais
Gullah

Hawaiian Creole English
Hiri Motu
Krio
Louisiana Creole
Mobilian Jargon
Palenquero
Papamientu
Pidgins and Creoles, Variation in
Russenorsk
Sango
Tok Pisin
Tsotsi Taal
Yanito

Sign Languages

Sign Languages of the World
Sign Language: Morphology

Sino-Tibetan Languages

Sinitic Languages
 Chinese
Tibeto-Burman Languages
 Burmese
 Karen Languages
 Tibetan

South American Languages

Andean Languages
 Aymará
 Quechua
Arawak Languages
 Tariana
Cariban Languages
Choco Languages
Macro-Jê
Mapudungan
Panoan Languages
Tucanoan Languages
Tupian Languages
 Guarani

Tai Kadai Languages

Lao
Thai

Uralic Languages

Estonian
Finnish
Hungarian
Nenets
Saami

Language Isolates and Languages of Disputed Affiliation

Ainu
 Basque
 Burushaski
 Elamite
 Etruscan
 Hurrian
 Japanese
 Ket
 Korean
 Nivkh
 Ryukyuan
 Sumerian
 Yukaghir

Artificial Languages

Esperanto

Issues in Language Typology and Classification

Endangered Languages

Ethnologue

Morphological Typology

Diachronic Morphological Typology

Arabic as an Introflecting Language

Chinese as an Isolating Language

Central Siberian Yupik as a Polysynthetic Language

Finnish as an Agglutinating Language

Italian as a Fusional Language

Long-Range Comparison: Methodological Disputes

Austic hypothesis

Austro-Tai hypothesis

Nostratic hypothesis

This page intentionally left blank

INTRODUCTION

In this volume, the world's leading experts describe many of the languages of the world. It is estimated that there are more than 250 established language families in the world, and over 6800 distinct languages, many of which are threatened or endangered. This volume provides the most comprehensive survey available on a large proportion of these. It contains 377 articles on specific languages or language families drawn from the two editions of the *Encyclopedia of Language and Linguistics* (ELL). The articles describe the sounds, meaning, structure, and family relationships of the languages, and have been chosen to illustrate the range and diversity of human language.

The *Concise Encyclopedia of Languages of the World* is unrivalled in its scope and content. We include articles on all the large language families, such as *Austronesian* by Tony Crowley, *Niger-Congo* by John Bendor-Samuel, and *Indo-European* by Neville Collinge; on many smaller families, like the North American *Iroquoian* by Marianne Mithun and *Caddoan* by David Rood; and on many 'language isolates', languages with disputed genetic affiliation to any other language, such as *Burushaski* by Greg Anderson, *Basque* by José Hualde, and *Japanese* by Masayoshi Shibatani. We have included a few languages which are no longer spoken but which have been important for historical linguistics, like *Ancient Egyptian* by John Ray, *Hittite* by J G McQueen, and *Pictish* by William Nicolaisen. There are also articles on pidgins and creoles spoken all over the world, from an article by Suzanne Romaine on *Tok Pisin* in Papua New Guinea to another by Raj Mesthrie on *Fanagalo* in southern Africa; as well as various articles on *Sign languages* by Wendy Sandler, Ulrike Zeshan, and Trevor Johnston respectively.

All the world's major languages are covered with articles on *Chinese* by Yueguo Gu, *Arabic* by Stephan Procházka, *Hindi* by Shaligram Shukla, and *Spanish* by Roger Wright. *English* is thoroughly described with articles on all its periods by Cynthia Allen (*Old English*), Jeremy J Smith (*Middle English*), Helena Raumolin-Brunberg (*Early Modern English*), Joan Beal (*Later Modern English*), Michael Swan (*English in the Present Day*), and Braj Kachru (*World Englishes*). Inevitably some of the languages described in this volume have very small numbers of speakers and hence are in danger of being overwhelmed and lost altogether. Some linguists estimate that as many as 50–80% of the world's languages may be at risk of extinction in the next century. Many communities and linguists around the world are working together to develop innovative ways of passing on their languages to future generations. The article *Endangered Languages* by Lenore Grenoble describes some of the reasons for language loss and proposes practical means of assessing language vitality.

The *Concise Encyclopedia of Languages of the World* is the definitive resource on the languages of the world in one compact volume. Each language article gives a brief description of the language and its speakers, together with any known or hypothesized genetic relationships, and highlights interesting phonological, semantic, and syntactic features. Similarly, the articles on language families outline the membership and distribution of the family and highlight any particular phonological, semantic, or syntactic features common to the family. There is a list of useful references for further reading at the end of each article. The articles are ordered alphabetically by language, so the reader who wishes to see the overall coverage in a particular family or area will find it helpful to consult the subject classification in the front of the volume. Many languages are known in the literature under different names or spellings. Authors have highlighted these differences, and, in some cases, explained why they have chosen one name or spelling over another. For ease of reference, all variant language

names and spellings are listed in the index. Just because a language does not have its own article, does not mean that it is not discussed in another article, so users of this volume are encouraged to work from the index in order to find information on the language they want.

The Notion 'Language'

The identification of different languages is not a straightforward matter. Every language is characterized by variation within the speech community that uses it. If the resulting speech varieties are sufficiently similar as to be considered merely characteristic of a particular geographic region or social grouping they are generally referred to as dialects, so Cockney and Norfolk are usually considered to be dialects of English. Sometimes social, political and historical pressures are such that the varieties are considered to be distinct enough to be treated as separate languages, like *Swedish* and *Norwegian* or *Hindi* and *Urdu*. Often the question of whether two languages are varieties of a single language or distinct languages is much argued over, like *Macedonian* and *Bulgarian*, or *English* and *Scots*. The naming of a language is another point of possible contention. While most linguists estimate around 6800 languages in the world, they also recognise four or five times that number of language names. A particular language may be known by one name to scholarship and another to its speakers; thus the name 'Akan' is not generally used by speakers of the language since Akan speech forms constitute a dialect continuum running from north to south in Ghana and different communities refer to their tongue by different names – *Asante*, *Fante*, *Twi*, *Akuapem*, *Brong*, *Akyem* or *Kwahu*.

Language Classification

Languages can be classified in a number of different ways and for a number of different purposes. The most common classification is 'genetic', which classifies languages into families on the basis of descent from a presumed common ancestor. 'Areal' classification groups languages together either on the basis of structural features shared across language boundaries within a geographical area, or more straightforwardly simply within a geographical area. A 'lexicostatistic' classification uses word comparisons as evidence of language relationships. A 'typological' classification supposes a small set of language types, traditionally word types (isolating, agglutinating, fusional, polysynthetic), to which languages can be assigned.

Genetic classification The article *Classification of Languages* by Barry Blake describes the principles underlying the classification of languages adopted in ELL2 and hence in this work. It is accompanied by a map showing the location of major language groupings worldwide. This approach is one in which languages are classified into families, based on divergence from a presumed common ancestor. Good examples are the *Dravidian languages* of Southern India and *Indo-European*. The Indo-European family includes most of the languages of Europe, Iran, Afghanistan, and the northern part of South Asia. These languages can be shown to descend from a common ancestor, a common protolanguage. There are no records of the ancestral language, but it can be reconstructed from records of daughter languages such as *Sanskrit*, *Ancient Greek*, and *Latin* by using what is known as the 'comparative method'. The method is briefly explained in the article. The comparative method relies on the existence of historical records and while this is possible for Indo-European and Dravidian languages, it is not possible in the same way for other proposed language families – the indigenous languages of the Americas or of Australia for example.

More speculative classifications, far from universally accepted, relate more language families together and hence try to explore language further back in time. These efforts are discussed in Lyle Campbell's article *Long-Range Comparison: Methodological Disputes*. One of the boldest and most controversial is the *Nostratic hypothesis*, which proposes a macrofamily consisting of Indo-European, Semitic, Berber, Kartvelian, Uralic, Altaic, Korean, Japanese, and Dravidian. Similarly ambitious is the proposed *Austro-Tai hypothesis* combining Hmong-Mien (Miao-Yao), the Tai-Kadai (or Daic) family, and Austronesian. The *Austroic hypothesis* extends this proposal to include Austroasiatic.

Areal classification There is a broader and a looser sense in which an areal classification can be useful. The looser sense simply groups languages together regionally. Here genetic affiliations are not firmly established but shared lexicon and similar structural features suggest that the languages in question have been in contact with each other over a long period of time. In the stricter sense, areal linguistics is concerned with the diffusion of structural features across language boundaries within a geographical area. The term 'linguistic area' refers to a geographical area in which, due to borrowing and language contact, languages of a region come to share certain structural features – not just loanwords, but also shared phonological, morphological, syntactic, and other

traits. The central feature of a linguistic area is the existence of structural similarities shared among languages where some of the languages are genetically unrelated, like Turkish and Greek in the Balkans. It is assumed that the reason the languages of the area share these traits is through contact and borrowing. In addition to a general article on *Areal Linguistics* by Lyle Campbell, this volume also includes articles on areas which have been particularly studied from an areal point of view: *Africa as a Linguistic Area* by Bernd Heine; *Balkans as a Linguistic Area* by Victor Friedman; *Ethiopia as a Linguistic Area* by Joachim Crass; *Europe as a Linguistic Area* by Thomas Stolz; *South Asia as a Linguistic Area* by Karen Ebert; *Southeast Asia as a Linguistic Area* by Walter Bisang.

Lexicostatistic classification Word comparisons were thought for a long time to be evidence of language family relationship, but, given a small collection of likely-looking words, it is difficult to determine whether they are really the residue of common origin and not due to chance or some other factor. Lexical comparisons by themselves are seldom convincing without additional support from other criteria. Most scholars require that basic vocabulary be part of the supporting evidence for any distant genetic relationship. Basic vocabulary is generally understood to include terms for body parts, close kinship, frequently encountered aspects of the natural world (mountain, river, cloud), and low numbers. Basic vocabulary is generally resistant to borrowing, so comparisons involving basic vocabulary items are less likely to be due to diffusion and stand a better chance of being inherited from a common ancestor than other kinds of vocabulary. Still, basic vocabulary can also be borrowed – though infrequently – so that its role as a safeguard against borrowing is not foolproof. Lexicostatistics are often used as partial evidence in discussing relationships between Southern American and African languages where there are few historical records: see for example the articles by Constenla Umaña on *Misumalpan* and *Chibchan*, and the article by David Dwyer on *Mande*.

Typological classification At the beginning of the nineteenth century, morphological studies identified a small set of language types related primarily to word structure. The main types were isolating (words are monomorphemic and invariable, as explained in the article on *Chinese as an Isolating Language* by Jerome Packard) agglutinating (words are formed by a root and a clearly detachable sequence of affixes, each of them expressing a separate item of meaning, as exemplified in the article *Finnish as an Agglutinating Language* by Fred Karlsson), fusional (words are formed by a root and (one or more) inflectional affixes, which are employed as a primary means to indicate the grammatical function of the words in the language; see *Italian as a Fusional Language* by Claudio Iacobini) and polysynthetic (the base is the lexical core of the word; it can be followed by a number of postbases e.g. *Central Siberian Yupik as a Polysynthetic Language* by Willem de Reuse). Further types have been added as explained in *Arabic as an Introflecting Language* by Janet Watson. This morphological typology is still of some relevance but with advances in grammatical and semantic description typological classification is nowadays refined. It extends to a range of other linguistic features and to an interest in ‘universal’ linguistic properties. Syntactic features such as word order differences between languages, case marking systems, tense and aspect distinctions, modal markers, for instance evidentiality, and serial verb construction. Phonological features such as consonant types, like ejectives or clicks, vowel or nasal harmony and stressmarking. It also includes discourse phenomena including topic marking, reference chaining, and switch reference. Features like these can be found in the index.

The articles in this volume provide fascinating insights into the structure, history, and development of language families and individual languages. They highlight the diversity of the world’s languages, from the thriving to the endangered and extinct. No other single volume matches the coverage of languages or the authority of the contributors of the *Concise Encyclopedia of Languages of the World*.

Keith Brown and Sarah Ogilvie

This page intentionally left blank

CONTRIBUTORS

C Abbott

University of Wisconsin, Green Bay, WI, USA

F Ackerman

University of California, San Diego, CA, USA

W F H Adelaar

Leiden University, Leiden, The Netherlands

T A Åfarli

Norwegian University of Science and Technology,
Trondheim, Norway

D Aguirre Licht

Universidad de los Andes, Bogotá, Colombia

A Y Aikhenvald

La Trobe University, Bundoora, VIC, Australia

C L Allen

Australian National University, Canberra, Australia

E Al-Wer

University of Essex, Wivenhoe Park, UK

A Amha

Leiden University, Leiden, The Netherlands

G D S Anderson

Max Planck Institute, Leipzig, Germany and University of
Oregon, Eugene, OR, USA

D Appleyard

University of London, London, UK

A Archi

University of Rome, Rome, Italy

I W Arka

Australian National University, Denpasar Bali,
Indonesia

J Augerot

University of Washington, Seattle, WA, USA

P K Austin

School of Oriental and African Studies, London, UK

P Bakker

Aarhus University, Aarhus, Denmark

Z Bao

National University of Singapore, Singapore

M Baptista

University of Georgia, Athens, GA, USA

J Barnes

SIL International, Bogota, Colombia

P O Bartlett

Vienna, VA, USA

W Bauer

Wellington, New Zealand

J Beal

University of Sheffield, Sheffield, UK

J Bendor-Samuel

Summer Institute of Linguistics, High Wycombe, UK

S B Benor

Hebrew Union College – Jewish Institute of Religion,
Los Angeles, CA, USA

A Berge

University of Alaska, Fairbanks, AK, USA

J Beswick

University of Southampton, Southampton, UK

P Bhaskararao

Tokyo University of Foreign Studies, Tokyo, Japan

T K Bhatia

Syracuse University, Syracuse, NY, USA

T Biberauer

Newnham College, Cambridge, UK

W Bisang

Johannes Gutenberg University, Mainz, Germany

B J Blake

La Trobe University, Bundoora, VIC, Australia

K M Booker

The University of Kansas, Lawrence, KS, USA

K Börjars

The University of Manchester, Manchester, UK

S Brainard

Summer Institute of Linguistics, Philippines, Manila, Philippines

J O Breedveld

Leiden University, Leiden, The Netherlands

G Breen

Institute for Aboriginal Development, Alice Springs, NT, Australia

G A Broadwell

State University of New York, Albany, NY, USA

J L Brockington

University of Edinburgh, Edinburgh, UK

L Campbell

University of Utah, Salt Lake City, UT, USA

U Canger

University of Copenhagen, Copenhagen, Denmark

S Chaker

INALCO, Paris, France

P Ni Chatháin

University College Dublin, Dublin, Ireland

J Cole

University of Illinois at Urbana-Champaign, Urbana, IL, USA

N E Collinge

Cambridge, UK

B Connell

York University, Toronto, ON, Canada

A Constenla Umaña

University of Costa Rica, San José, Costa Rica

J Crass

Johannes Gutenberg University, Mainz, Germany

J Cremona**T Crowley[†]**

University of Waikato, Hamilton, New Zealand

G Cunningham

University of Oxford, Oxford, UK

G Darms

University of Fribourg, Fribourg, Switzerland

A T de Castilho**W J de Reuse**

University of North Texas, Denton, TX, USA

B Demiraj

University of Munich, Munich, Germany

P Denwood

University of London, London, UK

G Deutscher

Leiden University, Leiden, The Netherlands

G J Dimmendaal

University of Cologne, Cologne, Germany

A Dolgopolsky

University of Haifa, Haifa, Israel

M Donohue

National University of Singapore, Singapore

E Drechsel

University of Hawai'i at Manoa, Honolulu, HI, USA

H Dua

Central Institute of Indian Languages, Mysore, India

J Le Dù

Université de Bretagne Occidentale, Brest, France

M Durrell

University of Manchester, Manchester, UK

P N Dutta Baruah

Central Institute of Indian Languages, Mysore, India

D Dwyer

Michigan State University, East Lansing, MI, USA

K Ebert

University of Zurich, Zurich, Switzerland

J Edmondson

University of Texas at Arlington, Arlington, TX, USA

J F Elwolde

Valladolid, Spain

G Escure

University of Minnesota, Minneapolis, MN, USA

A Esposito

Oxford English Dictionary, Oxford, UK

J Essegbey

University of Florida, Gainesville, FL, USA

N Evans

University of Melbourne, Melbourne, Australia

T Eythórsson

University of Iceland, Reykjavík, Iceland

T T Field

University of Maryland, Baltimore, MD, USA

D W Fleck

La Trobe University, Bundoora, VIC, Australia

M Florey

Monash University, Clayton, VIC, Australia

[†]Deceased.

C Fowler

University of Nevada, Reno, Nevada

V A Friedman

University of Chicago, Chicago, IL, USA

N Gabas Jr

Belém, Brazil

J W Gair

Cornell University, New York, NY, USA

P Geraghty

University of the South Pacific, Suva, Fiji

D Gil

Max Planck Institute for Evolutionary Anthropology,
Leipzig, Germany

C Githiora

University of London, London, UK

C Goddard

University of New England, Armidale, Australia

B Gopinathan Nair

St Xavier's College, Thiruvananthapuram, Kerala State,
India

R G Gordon Jr

SIL International, Dallas, TX, USA

R Graczyk

St. Charles Mission, Pryor, MT, USA

G B Gragg

University of Chicago, Chicago, IL, USA

J H Grayson

University of Sheffield, Sheffield, UK

M L Greenberg

University of Kansas, Lawrence, KS, USA

L Grenoble

Dartmouth College, Hanover, NH, USA

A L Grünberg

St Petersburg Institute of Linguistics, St Petersburg,
Russia

Y Gu

Chinese Academy of Social Sciences, Beijing,
China

W H G Haacke

University of Namibia, Windhoek, Namibia

J A Hackett

Harvard University, Cambridge, MA, USA

E P Hamp

University of Chicago, Chicago, IL, USA

R T Harms**A Harris**

State University of New York, Stony Brook,
NY, USA

J B Haviland

Reed College, Portland, OR, USA

E C Hawkesworth

University College London, London, UK

R J Hayward**B Heine**

Institut für Afrikanistik, Universität zu Köln, Köln,
Germany

B G Hewitt

SOAS, London, UK

J H Hill

University of Arizona, Tucson, AZ, USA

K C Hill

University of Arizona, Tucson, AZ, USA

R Hoogenraad

Alice Springs, NT, Australia

T Hopkins

Florida International University, Miami, FL, USA

J I Hualde

University of Illinois at Urbana-Champaign, Urbana,
IL, USA

T J Hudak

Arizona State University, Tempe, AZ, USA

G Hudson

Michigan State University, East Lansing, MI, USA

M Hutt

University of London, London, UK

C Hyslop

La Trobe University, Bundoora, VIC, Australia

C Iacobini

Università di Salerno, Salerno, Italy

B Ingham

SOAS, London, UK

P J Jaggard

University of London, London, UK

E H Jahr

Agder University, Kristiansand, Norway

L A Janda

University of North Carolina, Chapel Hill, NC, USA

J A Janhunen

University of Helsinki, Helsinki, Finland

D Jauncey

Australian National University, Canberra, Australia

M Job

University of Göttingen, Göttingen, Germany

L Johanson

Johannes Gutenberg University, Mainz, Germany

T Johnston

Macquarie University, Sydney, Australia

M C Jones

University of Cambridge, Cambridge, UK

B D Joseph

Ohio State University, Columbus, OH, USA

B Kachru

University of Illinois at Urbana-Champaign, Urbana, IL, USA

L D Kaplan

University of Alaska, Fairbanks, AK, USA

F Karlsson

University of Helsinki, Helsinki, Finland

F Katamba

Lancaster University, Lancaster, UK

T Kaufman

University of Pittsburgh, Forest Hills, PA, USA

R Kikusawa

National Museum of Ethnology, Osaka, Japan

R Kim

University of Pennsylvania, Philadelphia, PA, USA

A Kimenyi

California State University at Sacramento, Sacramento, CA, USA

G Kiraz

Beth Mardutho: The Syriac Institute, Piscataway, NJ, USA

K É Kiss

Hungarian Academy of Sciences, Budapest, Hungary

H Kitano

Aichi University of Education, Aichi, Japan

T Klingler

Tulane University, New Orleans, LA, USA

H Koch

Australian National University, Canberra, Australia

M E Kropp Dakubu

University of Ghana, Legon, Accra, Ghana

P Y L Laki

La Trobe University, Bundoora, VIC, Australia

H Landar**R J LaPolla**

La Trobe University, Bundoora, VIC, Australia

J R Lee

Summer Institute of Linguistics, Darwin, NT, Australia

J Leer**A L Lepschy**

University College London, London, UK

G Lepschy

University College London, London, UK

D Levey

Universidad de Cádiz, Cádiz, Spain

J W Lobel

University of Hawai'i, Manoa, Honolulu, HI, USA

T Lundskaer-Nielsen

University College London, London, UK

C Macafee

University of Aberdeen, Aberdeen, UK

C J MacKay

Ball State University, Muncie, IN, USA

K MacKinnon

University of Herefordshire, Black Isle, UK

J G Macqueen

University of Bristol, Bristol, UK

C M MacRobert

Oxford University, Oxford, UK

A Marcantonio

University of Rome 'La Sapienza,' Rome, Italy

L Marten

School of Oriental and African Studies, London, UK

Y Matras

University of Manchester, Manchester, UK

J M Maxwell

Tulane University, New Orleans, LA, USA

F Mc Laughlin

University of Florida, Gainesville, FL, USA

S Mchombo

University of California, Berkeley, CA, USA

S McLendon

City University of New York, NY, USA

S Meira

Leiden University, Leiden, The Netherlands

R Mesthrie

University of Cape Town, Cape Town, South Africa

A Mettouchi

University of Nantes, Nantes, France

B B Mfenyana

Kagiso, South Africa

J Miller

University of Auckland, Auckland, New Zealand

M Minegishi

Tokyo University of Foreign Studies, Tokyo, Japan

M Mithun

University of California, Santa Barbara, CA, USA

M J Mixco

University of Utah, Salt Lake City, UT, USA

M Miyashita

University of Montana, Missoula, MT, USA

S Munshi

University of Texas at Austin, Austin, TX, USA

I Nedjalkov

Institute of Linguistic Research, St Petersburg, Russia

G Newton

University of Sheffield, Sheffield, UK

W Nicolaisen

University of Aberdeen, Aberdeen, UK

R Nordlinger

University of Melbourne, Melbourne, Australia

B Nothofer

Universität Frankfurt, Frankfurt, Germany

D Nurse

Memorial University of Newfoundland, St. John's, NL, Canada

C Ó Baoill

University of Aberdeen, Aberdeen, UK

D Odden

The Ohio State University, Columbus, OH, USA

S Ogilvie

University of Oxford, Oxford, UK

K Owolabi

University of Ibadan, Ibadan, Nigeria

J L Packard

University of Illinois at Urbana-Champaign, Urbana, IL, USA

P L Patrick

University of Essex, Colchester, UK

A Pawley

Australian National University, Canberra, Australia

P J Payton

University of Exeter, Cornwall, UK

D H Pentland

University of Manitoba, Winnipeg, MB, Canada

J R Perry

University of Chicago, Chicago, IL, USA

J Powell

Royal Holloway, University of London, Egham, Surrey, UK

S Procházka

University of Vienna, Vienna, Austria

R L Rankin

University of Kansas, Lawrence, KS, USA

M Ratliff

Wayne State University, Detroit, MI, USA

H Raumolin-Brunberg

University of Helsinki, Helsinki, Finland

J D Ray

University of Cambridge, Cambridge, UK

G Reesink

Leiden University, Leiden, The Netherlands

L A Reid

University of Hawai'i, Honolulu, HI, USA

N Reid

University of New England, Armidale, Australia

R A Rhodes

University of California, Berkeley, CA, USA

E R Ribeiro

Museu Antropológico, Universidade Federal de Goiás, Goiás, Brazil

D Ridgeway

University of Edinburgh, Edinburgh, UK

S Romaine

Oxford University, Oxford, UK

D Rood

University of Colorado, Boulder, CO, USA

M Ross

The Australian National University, Canberra, Australia

R A Rothstein

University of Massachusetts, Amherst, MA, USA

P Rowlett

University of Salford, Salford, UK

C Rubino

Rockville, MD, USA

C Rudin

Wayne State College, Wayne, NE, USA

T Salminen

University of Helsinki, Helsinki, Finland

W J Samarin

University of Toronto, Toronto, Canada

P Sammallahti

University of Oulu, Oulu, Finland

T Sanchez

University of Pennsylvania, Philadelphia, PA, USA

W Sandler

University of Haifa, Haifa, Israel

T C Schadeberg**R Schäfer**

University of Göttingen, Göttingen, Germany

H F Schiffman

University of Pennsylvania, Philadelphia, PA, USA

A Schwegler

University of California, Irvine, CA, USA

D Septfonds

National Institute of Oriental Languages, Paris, France

F Serzisko

University of Cologne, Cologne, Germany

C Shackle

University of London, London, UK

D Sharma

King's College London, London, UK

K Shea

University of Kansas, Lawrence, KS, USA

M Shibatani

Rice University, Houston, TX, USA

S Shukla

Georgetown University, Washington, DC, USA

D Silverman

Nanvet, NY, USA

G F Simons

SIL International, Dallas, TX, USA

J M Y Simpson

University of Aberdeen, Aberdeen, UK

P O Skjærø

Harvard University, Cambridge, MA, USA

J C Smith

University of Oxford, Oxford, UK

J J Smith

University of Glasgow, Glasgow, UK

D A Smyth

SOAS, University of London, London, UK

D B Solnit

University of Michigan, Ann Arbor, MI, USA

W B Sperlich

Auckland, New Zealand

B Spolsky

Bar-Ilan University, Ramat Gan, Israel

A van der Spuy

University of the Witwatersrand, Johannesburg, South Africa

S N Sridhar**T Stolz**

Universität Bremen, Bremen, Germany

J Stonham

University of Newcastle, Newcastle upon Tyne, UK

P S Subrahmanyam

Annamalai University, Bangalore, India

M Swan

Didcot, Oxfordshire, UK

A Teffeteller

Concordia University, Montreal, Canada

A Terrill

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

P M K Thomas

University of Oxford, Oxford, UK

P W Thomas

School of Welsh, Cardiff University, Wales, UK

S Thomason

University of Michigan, Ann Arbor, MI, USA

H-R Thompson

School of Oriental and African Studies, London, UK

L Todd

University of Ulster, Coleraine, Northern Ireland

G H Toops

Wichita State University, Wichita, KS, USA

F R Trechsel

Ball State University, Muncie, IN, USA

S Trechter

California State University, Chico, CA, USA

E Tucker

University of Oxford, Oxford, UK

R Underhill

San Diego State University, San Diego, CA, USA

B Vaux

University of Wisconsin-Milwaukee, Milwaukee, WI, USA

M W Visser

University of Stellenbosch, Stellenbosch, South Africa

A Vovin

University of Hawaii at Manoa, Honolulu, HI, USA

B K Waghmar

University of London, London, UK

K Wali

Paris, France

D Ward

University of Edinburgh, Edinburgh, UK

J Watkins

School of Oriental and African Studies, London, UK

L J Watkins

Colorado College, Colorado Springs, CO, USA

J C E Watson

University of Oslo, Oslo, Norway

J Watt

Cardiff University, Cardiff, UK

J C Wells

University College London, London, UK

S Weninger

Philipps-Universität, Marburg, Germany

A Wetter

Johannes Gutenberg University, Mainz, Germany

M W Wheeler

University of Sussex, Brighton, UK

A V Williams

University of Manchester, Manchester, UK

K Williamson

University of Port Harcourt, Port Harcourt, Nigeria

H C Wolfart

University of Manitoba, Winnipeg, Manitoba, Canada

J U Wolff

Cornell University, Ithaca, NY, USA

R Wright

University of Liverpool, Liverpool, UK

Z Xu

State University of New York, Stony Brook, NY, USA

S Young

University of Maryland Baltimore County, Baltimore, MD, USA

U Zeshan

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

R D Zorc

Language Research Center, McNeil Technologies Inc., Hyattsville, MD, USA

G Zuckermann

University of Cambridge, Cambridge, UK

F Zúñiga

University of Zurich and University of Leipzig and Centro de Estudios Públicos, Zurich, Switzerland

This page intentionally left blank

LIST OF ABBREVIATIONS

| | |
|--------|---|
| A | act (in speech act theory); actor (tagmemics); addressee; agent; agentive; argument; author |
| ABESS | abessive |
| ABL | ablative |
| ABS | absolutive |
| ACC | accusative |
| ACT | active; actor |
| Ad | adjunct |
| ADESS | adessive |
| ADJ | adjective, -ival |
| AdjP | adjective phrase |
| ADV | adverb(ial) |
| AdvP | adverbial phrase |
| AFF | affective; affix |
| AFFIRM | affirmative |
| AGR | agreement |
| AGT | agent |
| AI | Artificial Intelligence |
| ALL | allative |
| AM | amplitude-modulated (signal) |
| Amer | American |
| AN | adjective precedes noun (in word order typology) |
| ANIM | animate |
| ANN | artificial neural network |
| ANT | anterior |
| ANTI | antipassive |
| AOR | aorist |
| AP | atomic phonology |
| APG | arc pair grammar |
| APPL | applicative |
| ART | article |
| ASCII | American Standard Code for Information Interchange |
| ASL | American Sign Language |
| ASP | aspect(ual) |
| ASR | automatic speech recognition |
| ASSOC | associative |
| ATN | augmented transition network |
| ATR | advanced tongue root (distinctive feature) |
| ATTR | attribute |

| | |
|--------------|--|
| Ausian | Australian Sign Language |
| AUX | auxiliary |
| <i>b.</i> | born |
| BASIC | Basic All-purpose Symbolic Instruction Code |
| BEN | benefactive |
| BEV | Black English Vernacular |
| BNC | British National Corpus |
| BSE | base-form |
| BSL | British Sign Language |
| C | clause; coda (of syllable); codomain (set theory); complement(izer); consonant |
| c-command | constituent command |
| c-structure | constituent structure |
| CA | componential analysis; contrastive analysis; conversation analysis |
| CALL | computer assisted language learning |
| CAP | control agreement principle |
| CAT | category; computer-assisted translation |
| CAUS | causative |
| CCG | combinatory categorial grammar |
| CD | communicative dynamism; conceptual dependency |
| CF | characteristic frequency; constant frequency |
| CFG | context-free grammar |
| CFL | context-free language |
| CFPSG | context-free phrase structure grammar |
| CG | categorial grammar |
| CL | computational linguistics |
| CLASS | classifier |
| CN | common noun |
| COLL | collective |
| COM | comitative |
| COMP | comparative; complement(izer) |
| CONJ | conjunction/conjugation |
| CONS | consonantal |
| CONT | continuant; continuative |
| COP | copula |
| COR | coronal |
| CP | complement(izer) phrase |
| cps | cycles per second |
| CS | context-sensitive |
| CSG | context-sensitive grammar |
| CV | consonant vowel structure/sequence |
| CV phonology | skeletal phonology |
| D-structure | deep structure |
| <i>d.</i> | died |
| DA | discourse analysis |
| DAF | delayed auditory feedback |
| DAG | directed acyclic graph |
| DAT | dative |
| DCG | definite clause grammar |
| DD | discourse domain |
| DDG | daughter dependency grammar |
| DECL | declarative |
| DEF | definite |
| DEM | demonstrative |
| DESID | desiderative |
| DEST | destinative |

| | |
|----------------|--|
| DET | determiner |
| DG | dependency grammar |
| DIM | diminutive |
| DIR | direction(al) |
| DIST | distributive |
| DM | discourse marker |
| DO | direct object |
| dp | determiner phrase |
| DRS | discourse representation structure |
| DRT | discourse representation theory |
| DS | deep structure; direct speech |
| DTR | daughter (in HPSG) |
| DU | dual |
| DYN | dynamic |
| EA | error analysis |
| EAP | English for academic purposes |
| ECM | exceptional case marking |
| ECP | empty category principle |
| EEG | electroencephalography |
| EFL | English as a foreign language |
| EL | elative |
| ELT | English Language Teaching |
| EMG | electromyograph(y) |
| EMPH | emphatic |
| ENCL | enclitic |
| Eng | English |
| equi | equi NP deletion (= identity erasure transformation) |
| ERG | ergative |
| ESL | English as a second language |
| ESP | English for Specific/Special Purposes |
| ESS | essive |
| EST | Extended Standard Theory |
| etym | etymology |
| EXCL | exclusive |
| EXIST | existential |
| EXP | experiencer |
| F | false (in truth table); formant |
| f-structure | functional structure |
| F ₀ | fundamental frequency |
| F ₁ | first formant |
| F ₂ | second formant |
| F ₃ | third formant |
| FACT | factive |
| FDS | free direct speech |
| FEM | feminine |
| FFP | foot feature principle |
| FG | functional grammar |
| fig. | figure |
| FIN | finite |
| FIS | free indirect speech |
| ff. | <i>foruit, flourished</i> , lived |
| FLA | first language acquisition |
| FM | frequency modulation |
| FSP | functional sentence perspective |
| FSTN | finite state transition network |

| | |
|--------------|--|
| FUT | future |
| FUG | functional unification grammar |
| GB | government and binding (theory) |
| GB-phonology | government-based phonology |
| GEN | gender; genitive |
| GER | gerund |
| GN | genitive precedes noun (in word order typology) |
| GPSG | generalized phrase structure grammar |
| GR | grammatical relation |
| GS | generative semantics |
| H | head (of construction); hearer/reader; high/superposed (code/variety, in adiglossic situation); high (pitch/tone) |
| HABIT | habitual |
| HCI | human-computer interaction |
| HFC | head feature convention |
| HFP | head feature principle |
| HG | head grammar |
| HON | honorific |
| HPSG | head-driven phrase structure grammar |
| HUM | human |
| HYPOTH | hypothetical |
| Hz | hertz |
| IA | Item-and-Arrangement [model of grammatical description] |
| IC | immediate constituent |
| I-E | Indo-European |
| IELTS | [British Council] International English Language Testing System |
| iff | if and only if |
| IGNOR | ignorative |
| IL | interlanguage |
| ILL | illative |
| IMP | imperative |
| IMPERS | impersonal |
| IMPERF | Imperfect(ive) |
| INAN | Inanimate |
| INCL | Including; inclusive |
| INCORP | Incorporating |
| INDEF | Indefinite |
| INDIC | Indicative |
| INF | infinitival; infinitive |
| INFL | Inflection |
| INSTR | Instrumental |
| INTERJ | Integration |
| INTERROG | interrogative |
| INTRANS | Intransitive |
| IO | indirect object |
| IP | inflection phrase; Item-and-process [model of grammatical description] |
| IPA | International Phonetic Alphabet |
| IR | inflectional rule; internal reconstruction |
| IRR | irrealis |
| IRREG | irregular |
| IS | indirect speech |
| ISA | subsumption/subclass 'is a' |
| IT | Information Technology |
| ITER | iterative |

| | |
|-------|--|
| K | set of situations (in speech act theory) |
| kHz | kilohertz |
| KWIC | keyword in context |
| L | language; low (pitch/tone); low/vernacular variety [in diglossia] |
| L1 | first language |
| L2 | second or foreign language |
| LAB | labial |
| LAD | language acquisition device |
| LARSP | language assessment, remediation, and screening procedure |
| LAT | lateral |
| LEX | lexicity (in HPSG) |
| LF | lexical function; logical form |
| LFG | Lexical Functional Grammar |
| lit. | literally |
| LMC | lower middle class |
| LOC | local; locative; locus |
| LP | language planning; linear precedence [statements]; linear prediction |
| LSP | language for special/specific purposes |
| LTAG | lexicalized tree adjoining grammar |
| LU | lexical unit |
| M | mid [tone]; Middle (in language names); modal |
| MASC | masculine |
| Mb | megabyte |
| MDS | multidimensional scaling |
| MG | Montague Grammar |
| MLAT | Modern Language Aptitude Test |
| MLU | mean length of utterance |
| MMC | middle-middle class |
| Mod | modern |
| MOD | modifier |
| MRI | magnetic resonance imaging |
| MT | mother tongue; machine translation |
| N | new (speaker); noun; nucleus (of syllable) |
| n.d. | no date |
| n.s. | new series |
| NA | noun precedes adjective (in word order typology) |
| NAS | nasal |
| NEG | negation; negative |
| NEUT | neuter |
| NG | noun precedes genitive (in word order typology) |
| NL | native language; natural language |
| NLG | natural language generation |
| NLP | natural language processing |
| NLU | natural language understanding |
| NMR | nuclear magnetic resonance |
| NN | neural net(work) |
| NNS | nonnative speaker |
| NOM | nominative; nominal(ization) |
| NP | noun phrase |
| NPrel | relative noun phrase |
| NRel | noun precedes relative clause (in word order typology) |
| NS | native speaker |
| nt | nonterminal |
| NT | New Testament |

| | |
|---------|--|
| NUM | number |
| NVC | non-verbal communication |
| O | onset (of syllable) |
| OBJ | object |
| OBL | oblique |
| OBS | obstruent |
| obs. | obsolete |
| OCR | optical character recognition |
| OED | <i>Oxford English Dictionary</i> |
| OOP | object-oriented programming |
| OPT | optative |
| OSV | object-subject-verb (in word order typology) |
| OT | Old Testament; Optimality Theory |
| OV | object precedes verb (in word order typology) |
| OVS | object-verb-subject (in word order typology) |
| P | phrase; predicate |
| PA | pushdown automation |
| PART | participle; particle; partitive |
| PASS | passive |
| PAT | patient |
| PERF | perfect(ive) |
| PERS | person(al) |
| PET | positron-emission tomography |
| PF | phonetic form (in principles and parameters framework) |
| PHON | phonology |
| PIE | Primitive Indo-European; Proto-Indo-European |
| PL | plural |
| PM | phrase marker |
| Po | postposition |
| PO | primary object |
| POL | polite |
| POSS | possessive; possessor |
| POTEN | potential |
| PP | prepositional phrase |
| pp | past participle |
| PLUPERF | pluperfect |
| PRED | predicative |
| PREF | prefix |
| PREP | preposition |
| PRES | present |
| PRO | an unspecified NP |
| pro | pronominal element; pronoun |
| PROG | progressive |
| ProgP | progressive phrase |
| PROHIB | prohibitive |
| PRESP | present participle |
| PS-rule | phrase structure rule |
| PSG | Phrase Structure Grammar |
| PTQ | [the] proper treatment of quantification [in English] (Montague grammar) |
| PURP | purpose; purposive |
| Q | question |
| QR | quantifier raising |
| QUANT | quantifier |
| QU | <i>wh</i> -marking |

| | |
|--------------|--|
| R-expression | referential/referring expression |
| R-graph | relational graph (in arc pair grammar) |
| RC | relative clause |
| RECIP | recipient/reciprocal |
| REFL | reflexive |
| reg | regular |
| ReIN | relative clause precedes noun (in sword order typology) |
| REP | repetitive |
| RES | resumptive/result |
| REST | Revised Extended Standard Theory |
| rev. | revised |
| RG | Relational Grammar |
| RNR | right node raising |
| RP | received pronunciation |
| RR | readjustment rule; redundancy rule |
| RST | Rhetorical Structure Theory |
| RT | reaction time; RTN recursive transition network |
| S | point of speech (temporal logic); sentence; sign (sign language); source; speaker; speaker/writer; standard (speaker); strong (syllable); subject (tagmemics); subject term (or conclusion in a syllogism) |
| S-structure | surface structure |
| SAE | Standard American English; standard average European OVhorO |
| SC | small clause; structural change |
| SD | structural description |
| SEM | semantics |
| SGML | standard generalized markup language |
| SIB | sibilant |
| sing | singular |
| SL | source language |
| SLA | second language acquisition |
| SLASH | unbounded dependency (in HPSG) |
| SON | sonorant |
| SOV | subject-object-verb (in word order typology) |
| SPEC | specifier |
| SS | surface structure |
| SSC | specified subject condition |
| Sta | statement |
| STAT | static |
| STRID | strident |
| SUBCAT | subcategorization |
| SUBJ | subject; subjunctive |
| SUBJUNC | subjunctive |
| SUBORD | subordinate, subordinative |
| SUF | suffix |
| SUP | supine |
| SUPERESS | superessive |
| SV | subject precedes verb (in word order typology) |
| SVO | subject-verb-object (in word order typology) |
| SYLL | syllabic; syllable |
| SYN | synonym; syntax |
| T | tense; text; time; transformation; tree; true (in truth table); tu (= familiar pronoun of address) |
| T | trace |
| T-rule | transformational rule |
| TAG | Tree-Adjoining Grammar |

| | |
|----------|--|
| TAL | tree-adjoining language |
| TBU | tone-bearing unit |
| TC | total communication [approach] (in schools for the deaf) |
| TEFL | Teaching English as a foreign language |
| TEMP | temporal |
| TERM | terminative |
| TESOL | Teaching of English to Speakers of Other Languages |
| TG | Transformational Grammar |
| TGG | Transformational Generative Grammar |
| TL | target language |
| TNS | tense |
| TOEFL | Test of English as a Foreign Language |
| TOP | topic(alization) |
| TRANS | transitive |
| TRANSLV | translative |
| TYP | type |
| U | utterance |
| UCG | Unification Categorical Grammar |
| UG | Universal Grammar |
| UMC | upper middle class |
| V | verb(al); vowel; “our (= polite pronoun of address) |
| V | short vowel |
| V | long vowel |
| V-form | honorific form (of address) |
| VFORM | verb form |
| VIS | visual |
| VLSI | very large scale integration |
| VN | verbal noun |
| VO | verb precedes object (in word order typology) |
| VOC | vocalic |
| VOS | verb-object-subject (in word order typology) |
| VOT | voice onset time |
| VP | verb phrase |
| VS | verb precedes subject (in word order typology) |
| VSO | verb-subject-object (in word order typology) |
| W | weak (syllable) |
| WF | word formation |
| WFF | well-formed formula |
| WG | word grammar |
| WH-word | question word (<i>what, which</i> , etc.) |
| WP | Word-Paradigm (grammar) |
| 0 | zero (covert element) |
| 1 | first person |
| α | alpha, a variable |
| Σ | sentence; superfoot (in metrical phonology) |
| σ | syllable |

A

Abkhaz

B G Hewitt, SOAS, London, UK

© 2006 Elsevier Ltd. All rights reserved.

The Abkhaz language (/[a.]'aps.(wa bəz.'i)f^wa/) belongs to the North West Caucasian family (*see Caucasian Languages*). Abkhazians traditionally occupied the triangle framed in northwestern Transcaucasia between the Black Sea, the Greater Caucasus, and the river Ingur; the river Psou is now the northern frontier. This territory comprises the Republic of Abkhazia (/a.ps.'nə/, capital Aq^w'a, aka Sukhum), *de facto* independent since the war with Georgia (1992–1993) but in international law, deemed to be part of Georgia still. For most of the Soviet period it was an autonomous republic.

A wave of migrants out of Abkhazia after the Mongol incursions (14th century) removed the most divergent dialect, T'ap'anta, to the northern Caucasus (Karachay-Cherkessia). Consolidated there by Ashkarywa dialect speakers (17th and 18th centuries), today's Abaza population descended from them. Following Russia's conquest of the northwest Caucasus in 1864, most North West Caucasian speakers (including the now extinct Ubykhs) migrated to Ottoman lands, where the diaspora-communities (predominantly in Turkey) vastly outnumber the homelander; even so, the surviving languages are endangered in all locations. The dialects of Sadz, Akhch'ypsy, and Ts'abal are no longer attested in Abkhazia; only northern Bzyp and southern Abzhywa remain. Of the 102 938 Soviet Abkhazians recorded in 1989, 93 267 resided in Abkhazia, constituting 17.8% of the population. The single largest ethnic group in Abkhazia in 1989 were the Mingrelians; Abazas totalled 33 801. Though 93.3% of Abkhazians claimed fluency in Abkhaz, younger generations tend to use Russian (or Turkish).

The 17th-century, half-Abkhazian traveller Evliya Çelebi provided the first linguistic evidence. P. Uslar produced the first grammar (1862–1863), devising a Cyrillic-based script. An adaptation of this alphabet served the Abkhazians when the Soviets assigned them literary status (1921), though two different

roman orthographies were tried during the infant USSR's *latinizatsija*-drive. A Georgian orthography was imposed in 1938 and replaced by another Cyrillic alphabet in 1954. This one is still used, albeit with a recent reform to regularize labialization-marking. Abaza acquired literary status only in 1932; the Abkhaz and Abaza Cyrillic scripts diverge markedly.

A comprehensive list of phonemes appears in Table 1.

Certain idiolects have /f/ only in /a.'f'a/ 'thin' (otherwise /a.'p'a/). Bzyp boasts 67 phonemes by adding /f_ɛ d_ɛ f_ɛ' ɛ z ɛ^w z^w/ to the alveolo-palatals and /χ^ɛ χ^{ɛw}/ to the back fricatives. A glottal stop, apart from possibly realizing intervocalic /q'/, is also heard in /ʔaj/ 'no' (cf., /a:j/ 'yes'). Open vowel /a/ contrasts with close /ə/; /a:/ might also be phonemic. Stress is distinctive.

Abkhaz(-Abaza) is unique among Caucasian languages in not employing case-markers for the verb's major arguments, relying purely on pronominal crossreferencing within the polysynthetic verb; this patterning with three sets of affixes confirms the family's ergative nature. Some preverbs distinguish directionality via an a-grade (essive/illative/allative)

Table 1 Consonantal phonemes for literary (Abzhywa) Abkhaz

| p | b | p' | f | v | m | w |
|--|--|---|---------------------|---------------------|---|---|
| t | d | t' | | | n | r |
| t ^w [tʰ] | d ^w [dʰ] | t' ^w [tʰ'] | s | z | | |
| f _ɛ | d _ɛ | f _ɛ ' | | | | |
| f _ɛ ^w [f _ɛ ʰ] | d _ɛ ^w [d _ɛ ʰ] | f _ɛ ^w ' [f _ɛ ʰ'] | | | | |
| tʃ | dʒ | tʃ' | ʃ | ʒ | | |
| | | | ʃ ^w [ʃʰ] | ʒ ^w [ʒʰ] | | |
| ʃ | ʒ | ʃ' | ʂ | ʐ | | |
| | | | | | | l |
| | | | | | | j |
| | | | | | | q |
| k | g | k' | | | | |
| k ^j | g ^j | k ^j ' | | | | |
| k ^w | g ^w | k ^w ' | | | | |
| | | q' | χ | ɣ | | |
| | | q ^j | χ ^j | ɣ ^j | | |
| | | q ^w | χ ^w | ɣ ^w | | |
| | | | h | | | |
| | | | h ^w [hʰ] | | | |

vs. a reduced/zero grade (relative/ablative) for the specified location.

The Stative-Dynamic opposition, verbal complexity, the relative strategy, the potential/involuntary constructions, and the preverbal grade-system are illustrated below:

- (1) a-p'h^wəs a-ma'q'a ø-'lə-mkə-w-p'
the-womanII the-beltI itI-sheII-wear-
Stat-Fin.Pres
'The woman is wearing the/a belt'
- (2) a-p'h^wəs a-ma'q'a
he-womanII the-beltI
ø-'lə-mkə-l-fs'a-ø-r.t^w
itI-herII-Prev-sheIII-put-Past.N/F.Aor-Res
ø-sə-z-'lə-r-q'a-fs'a- / ø-sə-'zə-q'a-fs'a-
wa-m wa-m
itI-II-Pot-herII-Caus- itI-II-Pot-Prev-do-
Prev-do-Dyn-not.Pres Dyn-not.Pres
'I cannot make the woman put on (herself/some
other woman) the belt'
- (3) a-p'h^wəs a-ma'q'a
he-womanII the-beltI
ø-'lə-mkə-l-χə-ø-r.t^w
itI-herII-Prev-sheIII-take-Past.N/F.Aor-Res
Ø-'s-amχa-lə-r-q'a- / Ø-'s-amχa-
fs'a-ø-jt' q'a-ŋp'a-ø-jt'
itI-II-unwilling-herII- itI-II-unwilling-
Caus-Prev-do-Past- Prev-do-Past-
Fin.Aor Fin.Aor
'I unwillingly/involuntarily got the woman
to remove the belt (from herself/some
other woman)'

- (4) a-ma'q'a ø-'zə-mkə-z-z-χə-ø-z a-p'h'wəs
the-beltI itI-whoII-Prev-whoIII- the-womanI
 take-Past-N/F.P/I
d-'sə-pj^wma-w-p'
sheI-myII-wife-Stat-Fin.Pres
'The woman who took off her belt is my wife'

The lexicon reveals Iranian, Turkish, Russian, and Kartvelian (mainly Mingrelian) influences.

Bibliography

- Allen W S (1956). 'Structure and system in the Abaza verbal complex.' In *Transactions of the Philological Society*. Oxford: Basil Blackwell. 127–176.
- Catford J C (1972). 'Labialisation in Caucasian languages, with special reference to Abkhaz.' In Rigault A & Charbonneau R (eds.) *Proceedings of the seventh international congress of phonetic sciences, 22–28 August 1971*. The Hague: Mouton. 679–681.
- Chirikba V A (2003). *Abkhaz: languages of the world/materials* 119. München/Newcastle: Lincom Europa.
- Dumézil G (1967). *Documents anatoliens sur les langues et les traditions du Caucase. V. Etudes Abkhaz*. Paris: Librairie Adrien-Maisonneuve.
- Hewitt B G (2005). *Abkhazian folktales (grammatical introduction, texts, translation, and vocabulary)*. München/Newcastle: Lincom Europa.
- Hewitt B G & Khiba Z K (1989). *Lingua descriptive studies 2: Abkhaz*. Chippenham: Routledge.
- Hewitt B G & Khiba Z (1998). *Abkhaz newspaper reader (with supplements)*. Kensington: Dunwoody Press (MRM).
- Spruit A (1986). 'Abkhaz Studies.' Ph.D. diss., Leiden.

Adamawa-Ubangi

J Bendor-Samuel, Summer Institute of Linguistics,
High Wycombe, UK

© 2006 Elsevier Ltd. All rights reserved.

The languages grouped together as Adamawa-Ubangi belong to the Volta-Congo branch of the Niger-Congo family. These languages are spoken across central Africa in an area that stretches from north-eastern Nigeria through northern Cameroon, southern Chad, the Central African Republic (CAR), and northern Zaire into southwestern Sudan.

The Speakers

In the absence of firm figures, the number of speakers of languages in this group can only be estimated at

around eight to nine million people. Several languages with a million or more speakers belong to this group (e.g., Zande in CAR, Zaire, and Sudan; Ngbaka in North Zaire; and Gbaya in CAR and Cameroon).

Study of the Group

Little study of the languages in this group was undertaken before the 20th century. Westermann and Bryan (1952) treated them as individual units or clusters. Greenberg (1963) was the first to group them together as a branch of Niger-Congo. He used the name 'Adamawa-Eastern' for this group of languages. Samarin (1971) suggested the use of the name 'Ubangi' to replace 'Eastern.' Boyd (1989) has summarized recent studies on this language group, showing that for many of the languages there has

been little detailed research. This is particularly true of the Adamawa languages. Knowledge of many of them is very sketchy.

Classification

The languages fall into two main groups – Adamawa and Ubangi. The Adamawa languages are found in northern Nigeria, Cameroon, and Chad, whereas the Ubangi languages are spoken in CAR, northern Zaire, and southwestern Sudan.

The Adamawa languages are divided into 16 groups: Waja (at least 6 languages), Leko (4 languages), Duru (18 languages), Mumuye (9 languages), Mbum (7 languages), Yungur (5 languages), Kam, Jen (2 languages), Longuda, Fali, Nimbari, Bua (9 languages), Kim, Day, Burak (6 languages), and Kwa.

Lexicostatistic studies show that the relationship among the groups is loose, but some of them can be grouped together so that two or perhaps three clusters emerge. The Leko, Duru, Mumuye, and Nimbari groups form a core of closely related languages. Another cluster comprises Mbum, Bua, Kim, and Day. Possibly a third cluster of Waja, Longuda, Yungur, and Jen can be formed.

The Ubangi languages show a much closer relationship to each other than do the Adamawa languages, and they fall into six main groups: Gbaya (4 languages), Banda, Ngbandi, Sere (6 languages), Ngbaka-Mba (9 languages), and Zande (5 languages).

Structural Features

Phonetics and Phonology

In Adamawa languages the set of initial consonants is much larger than the set of noninitial consonants,

whereas in Ubangi languages there is little difference in size between the two sets of consonants. Most languages have either a five- or seven-vowel system. Two, three, or four contrastive tones are found. Downstep is not common.

Grammar and Syntax

Noun class systems are not universal and are found mainly in the Adamawa languages. Some only comprise paired singular and plural suffixes without concord markers.

Verb systems usually contrast perfective and imperfective forms. Verbal extensions mark iteration, intensive, benefactive, and causative. Generally, inflectional morphemes are prefixed, and derivational morphemes are suffixed.

The predominant sentence word order is SVO. Negative markers occur clause final, and interrogative markers and words occur sentence final.

Bibliography

- Boyd R (1989). 'Adamawa-Ubangi.' In Bendor-Samuel J (ed.) *The Niger-Congo languages*. Lanham and London: University Press of America. 178–215.
- Greenberg JH (1963). *The languages of Africa*. The Hague: Mouton & Co.
- Kleinewillinghofer U (1996). 'Die nordwestlichen Adamawa-Sprachen.' *Frankfurter Afrikanistische Blätter* 8, 81–104.
- Samarin W J (1971). 'Adamawa-Eastern.' In Sebeok T A (ed.) *Current trends in linguistics, vol. 7: Linguistics in Sub-Saharan Africa*. The Hague: Mouton.
- Westermann D & Bryan M A (1952). 'Languages of West Africa.' In *Handbook of African Languages II*. London: Oxford University Press.

Africa as a Linguistic Area

B Heine, Institut für Afrikanistik, Universität zu Köln, Köln, Germany

© 2006 Elsevier Ltd. All rights reserved.

On Linguistic Areas

A number of different definitions of linguistic areas have been proposed; what is common to most of them are the following characteristics:

1. There are a number of languages spoken in one and the same general area.

2. The languages share a set of linguistic features whose presence can be explained with reference to neither genetic relationship, drift, universal constraints on language structure or language development, nor to chance.
3. This set of features is not found in languages outside the area.
4. On account of (2), the presence of these features must be the result of language contact.

Among the linguistic areas (or Sprachbunds) that have been proposed, perhaps the most widely

recognized are the Balkans and Meso-America. The African continent has been said to form a linguistic area, but so far there is no conclusive evidence to substantiate this statement.

Earlier Work

While there were a number of studies on areal relationship in Africa in the earlier history of African linguistics, Greenberg (1959) constitutes the first substantial contribution to this field. In an attempt to isolate areal patterns both within Africa and separating Africa from other regions of the world, he proposed a number of what he called 'special' features of African languages. The properties listed by Greenberg include in particular a number of lexical polysemies, such as the use of the same term for 'meat' and '(wild) animal,' the use of the same term for 'eat,' 'conquer,' 'capture a piece in a game,' and 'have sexual intercourse,' and the use of a noun for 'child' as a diminutive or of 'child of tree' to denote 'fruit of tree.' Another noteworthy contribution to areal relationship within Africa appeared in 1959: Larochette (1959) presented a catalog of linguistic properties characteristic of Congolese Bantu (Kikongo [Kituba], Luba, and Mongo [Mongo-Nkundu]), an Ubangi language (Zande), and a Central Sudanic language (Mangbetu), but many of the properties proposed by him can also be found in other regions and genetic groupings of Africa. A catalog of properties characterizing African languages was also proposed by Welmers (1974) and Gregersen (1977). Building on the work of Greenberg (1959) and Larochette (1959), Meeussen (1975) proposed an impressive list of what he called 'Africanisms,' that is, phonological, morphological, syntactic, and lexical properties widely found in African languages across genetic boundaries.

Another seminal publication on areal relationship was published by Greenberg in 1983. Noting that there are no areal characteristics found everywhere in Africa but nowhere else, he proceeded to define areal properties "as those which are either exclusive to Africa, though not found everywhere within it, or those which are especially common in Africa although not confined to that continent" (1983: 3). As an example of the former, he mentioned clicks; as instances of the latter, he discussed in some detail the following four properties: (1) coarticulated labiovelar stops, (2) labiodental flaps, (3) the use of a verb meaning 'to surpass' to express comparison, and (4) a single term meaning both 'meat' and '(wild) animal.' He demonstrated that these four properties occur across genetic boundaries and, hence, are

suggestive of being Pan-African traits, especially since they are rarely found outside Africa.

Greenberg (1983) went on to reconstruct the history of these properties by studying their genetic distribution. He hypothesized that (1), (3), and (4) are ultimately of Niger-Kordofanian origin, even though they are widely found in other African families, in particular in Nilo-Saharan languages. For (2), however, he did not find conclusive evidence for reconstruction, suggesting that it may not have had a single origin but rather that it arose in the area of the Central Sudanic languages of Nilo-Saharan and the Adamawa-Ubangi languages of Niger-Congo.

Search for areal properties across Africa is associated not the least with creole linguistics. In an attempt to establish whether, or to what extent, the European-based pidgins and creoles on both sides of the Atlantic Ocean have been shaped by African languages, students of creoles pointed out a number of properties that are of wider distribution in Africa, perhaps the most detailed study being Gilman (1986).

Pan-African Properties

The term 'Pan-African properties' refers to linguistic properties that are (1) common in Africa but clearly less common elsewhere, (2) found at least to some extent in all major geographical regions of Africa south of the Sahara, and (3) found in two or more of the four African language families. The following catalog of selected properties is based on previous work on this subject (especially Greenberg, 1959, 1983; Larochette, 1959; Meeussen, 1975; Gilman, 1986).

A general phonological property that has been pointed out by a number of students of African languages is the preponderance of open syllables and an avoidance of consonant clusters and diphthongs. Furthermore, tone as a distinctive unit is characteristic of the majority of African languages, in most cases on both the lexical and grammatical levels.

Ignoring click consonants, which are restricted to southern Africa and three languages in East Africa (Sandawe, Hadza, and Dahalo), there are a number of consonant types that are widespread in Africa but uncommon elsewhere. This applies among others to coarticulated labiovelar stops, (especially *kp* and *gb*), which occur mainly in a broad geographical belt from the western Atlantic to the Nile-Congo divide. Perhaps even more characteristic are labiodental flaps, produced by the lower lip striking the upper teeth; although restricted to relatively few languages, they are found in all families except Khoisaan. A third type of consonants that is widespread in Africa but

rarely found outside Africa can be seen in voiced implosive stops.

In their arrangement of words, African languages of all four families exhibit a number of general characteristics such as the following: While on a worldwide level languages having a verb-final syntax (SOV) appear to be the most numerous, in Africa there is a preponderance of languages having subject-verb-object (SVO) as their basic order: Roughly 71% of all African languages exhibit this order. Furthermore, the placement of nominal modifiers after the head noun appears to be more widespread in Africa than in most other parts of the world. Thus, in Heine's (1976: 23) sample of 300 African languages, demonstrative attributes are placed after the noun in 85%, adjectives in 88%, and numerals in 91% of all languages.

Logophoric marking appears to constitute a specifically African construction type. Logophoric pronouns indicate coreference of a nominal in the nondirect quote to the speaker encoded in the accompanying quotative construction, as opposed to its noncoreference indicated by an unmarked pronominal device (concerning the areal distribution of these pronouns, see Güldemann, 2003).

Perhaps the most conspicuous area where one might expect to find Pan-African properties can be seen in lexical and grammatical polysemies. A number of examples of polysemy, such as 'meat'/'animal,' 'eat'/'conquer,' and so on, were mentioned earlier. Furthermore, there are some grammaticalization processes that are common in Africa but rare elsewhere, examples being the grammaticalization of body parts for 'stomach/belly' to spatial concepts for 'in(side),' or of verbs meaning 'surpass,' 'defeat,' or 'pass' to a standard marker of comparison (Heine, 1997: 126–129).

Quantitative Evidence

Being aware that for many of the Pan-African properties that have been discussed in the relevant literature there is only sketchy cross-linguistic information, Heine and Zelealem (2003) use a quantitative approach to determine whether Africa can be defined as a linguistic area. For each of the 149 languages of their sample, of which 99 are African languages and 50 are languages from other continents, they apply 11 criteria that have figured in previous discussions on the areal status of African languages. The criteria and main results of their African survey are listed in Table 1, and those of their worldwide sample in Table 2. What Table 2 suggests is the following:

Table 1 Relative frequency of occurrence of 11 typological properties in African languages^a

| <i>Properties used as criteria</i> | <i>Number of languages having that property</i> | <i>Percentage of all languages</i> |
|--|---|------------------------------------|
| 1. Labiovelar stops | 39 | 39.4 |
| 2. Implosive stops | 36 | 36.4 |
| 3. Lexical and/or grammatical tones | 80 | 80.8 |
| 4. ATR-based vowel harmony | 39 | 39.4 |
| 5. Verbal derivational suffixes (passive, causative, benefactive, etc.) | 76 | 76.7 |
| 6. Nominal modifiers follow the noun | 89 | 89.9 |
| 7. Semantic polysemy 'drink/pull, smoke' | 74 | 74.7 |
| 8. Semantic polysemy 'hear/see, understand' | 72 | 72.7 |
| 9. Semantic polysemy 'animal, meat' | 40 | 40.4 |
| 10. Comparative constructions based on the schema [X is big defeats/surpasses/ passes Y] | 82 | 82.8 |
| 11. Noun 'child' used productively to express diminutive meaning | 50 | 50.5 |

^aSample: 99 languages. Parameters 3, 7, and 8 have two options; if one of the options applies, this is taken as positive evidence that the relevant property is present.

Table 2 Distribution of 11 typological properties according to major world regions^a

| <i>Region</i> | <i>Total of languages</i> | <i>Total of properties</i> | <i>Average number of properties per language</i> |
|---------------------|---------------------------|----------------------------|--|
| Europe | 10 | 11 | 1.1 |
| Asia | 8 | 21 | 2.6 |
| Australia/Oceania | 12 | 37 | 3.0 |
| The Americas | 14 | 48 | 3.4 |
| Africa | 99 | 669 | 6.8 |
| Pidgins and creoles | 6 | 14 | 2.3 |
| <i>All regions</i> | <i>149</i> | | |

^aSample: 99 African and 50 non-African languages.

1. Africa clearly stands out against other regions of the world in having on average 6.8 of the 11 properties, while in other regions clearly lower figures are found.

6 Africa as a Linguistic Area

2. Outside Africa, no language has been found to have as many as five properties, while African languages have between 5 and 10 properties.

Isopleth Mapping

To study the internal structure of linguistic areas, isopleth mapping has been employed in linguistic areas such as South Asia (Masica, 1976), the Balkans (van der Auwera, 1998), and Meso-America (van der Auwera, 1998). Isopleth maps are designed on the basis of the relative number of features that languages of a linguistic area share: languages having the same number of properties, irrespective of which these properties are, are assigned to the same isopleth and, depending on how many properties are found in a given language, the relative position of that language within the linguistic area can be determined.

Applying isopleth mapping to Africa yields the following results: The most inclusive languages, having nine or more properties, are found in West Africa, including both Niger-Congo and Afro-Asiatic languages. A secondary isopleth center is found in the Cameroon–Central Africa area, where up to nine

properties are found. Clearly less central are languages farther to the west and south, that is, Atlantic and Mande languages on the one hand, and Bantu languages on the other, where around six properties are found. Peripheral Africa consists of the Ethiopian Highlands (*see Ethiopia as a Linguistic Area*) and northern (Berber) Africa, where less than five properties are found. **Figure 1** is based on an attempt to reduce the complex quantitative data to an isopleth map.

Conclusion

While there is no linguistic property that is common to all of the 2000-plus African languages, it seems possible on the basis of the quantitative data presented to define Africa as a linguistic area: African languages exhibit significantly more of the 11 properties listed in **Table 1** than non-African languages do, and it is possible to predict with a high degree of probability that if there is some language that possesses more than five of these 11 properties, then this must be an African language. Not all of the properties, however, are characteristic of Africa only; some are equally common in other parts of the world.

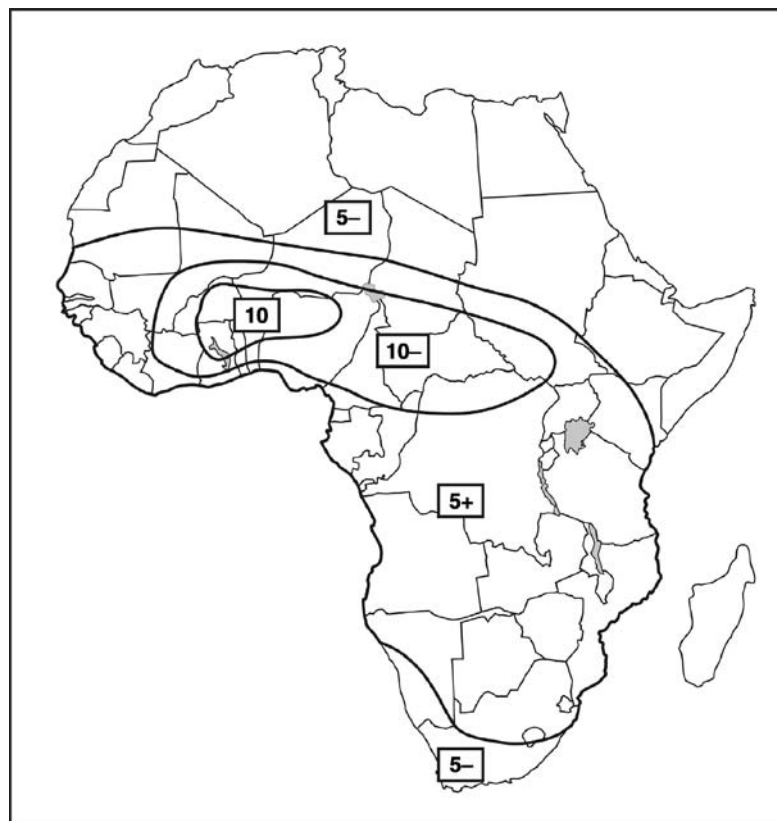


Figure 1 An isopleth sketch map of Africa based on 11 properties (sample: 99 languages).

Bibliography

- Aikhenvald A Y & Dixon R M W (eds.) (2001). *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press.
- Campbell L, Kaufman T & Smith-Stark T C (1986). 'Meso-America as a linguistic area.' *Language* 62(3), 530–570.
- Dimmendaal G J (2001). 'Areal diffusion versus genetic inheritance: an African perspective.' In Aikhenvald A Y & Dixon R M W (eds.) *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press. 358–392.
- Ferguson C A (1976). 'The Ethiopian language area.' In Bender M L, Bowen J D, Cooper R L & Ferguson C A (eds.) *Language in Ethiopia*. London: Oxford University Press. 63–76.
- Gilman C (1986). 'African areal characteristics: Sprachbund, not substrate?' *Journal of Pidgin and Creole Languages* 1(1), 33–50.
- Greenberg J H (1959). 'Africa as a linguistic area.' In Bascom W R & Herskovits M J (eds.) *Continuity and change in African Cultures*. Chicago: University of Chicago. 15–27.
- Greenberg J H (1963). *The languages of Africa*. The Hague: Mouton.
- Greenberg J H (1983). 'Some areal characteristics of African languages.' In Dihoff I R (ed.) *Current approaches to African linguistics*, vol. 1. Dordrecht: Foris Publications. 3–21.
- Gregersen E A (1977). *Language in Africa: an introductory survey*. New York, Paris, London: Gordon and Breach.
- Güldemann T (1997). *The Kalahari Basin as an object of areal typology: a first approach* (Khoisan Forum, 3). Cologne: Institut für Afrikanistik.
- Güldemann T (2003). 'Logophoricity in Africa: an attempt to explain and evaluate the significance of its modern distribution.' *Sprachtypologie und Universalienforschung (STUF)* 56(4), 366–387.
- Heine B (1975). 'Language typology and convergence areas in Africa.' *Linguistics* 144, 26–47.
- Heine B (1976). *A typology of African languages based on the order of meaningful elements* (Kölner Beiträge zur Afrikanistik, 4). Berlin: Dietrich Reimer.
- Heine B (1997). *Cognitive foundations of grammar*. Oxford, New York: Oxford University Press.
- Heine B & Kuteva T (2001). 'Convergence and divergence in the development of African languages.' In Aikhenvald A & Dixon R M W (eds.) *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press. 393–411.
- Heine B & Nurse D (eds.) (2000). *African languages: an introduction*. Cambridge: Cambridge University Press.
- Heine B & Zelealem L (2003). 'Comparative constructions in Africa: an areal dimension.' *Annual Publication in African Linguistics* 1, 47–68.
- Larochette J (1959). 'Overeenkomst tussen Mangbetu, Zande, en Bantu-talen.' *Handelingen van het XXIIIe Vlaams Filologencongres (Brussels) 1959*, 247–248.
- Masica C P (1976). *Defining a linguistic area: South Asia*. Chicago, London: University of Chicago Press.
- Meeussen A E (1975). 'Possible linguistic Africanisms.' Fifth Hans Wolff Memorial Lecture. Language Sciences, 35. Bloomington: Indiana University.
- Sandfeld K (1930). *Linguistique balkanique: problèmes et résultats* (Collection Linguistique, 31). Paris: Champion.
- Tosco M (2000). 'Is there an "Ethiopian language area"?' *Anthropological Linguistics* 42(3), 329–365.
- van der Auwera J (1998). 'Revisiting the Balkan and Meso-American linguistic areas.' *Language Sciences* 20(3), 259–270.
- Welmers W E (1974). *African language structures*. Berkeley: University of California Press.
- Westermann D (1935). 'Charakter und Einteilung der Sudansprachen.' *Africa* 8(2), 129–148.

Afrikaans

T Biberauer, Newnham College, Cambridge, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Afrikaans is the youngest fully standardized member of the West Germanic branch of the Indo-European language family. A daughter of Dutch (*Afrikaans* = the Dutch adjective meaning 'African'), it is primarily spoken in South Africa, where it is one of 11 official languages. Currently, it boasts the third largest speaker population, with only Zulu and Xhosa being more widely spoken (1996 Census). Afrikaans

also represents a minority language in Namibia and, increasingly, in expatriate communities, notably in Britain, Australia, New Zealand, and Canada.

History

The precise circumstances surrounding the development of Afrikaans as a language in its own right have been energetically disputed. What is uncontroversial is that the Dutch East India Company's establishment of a refreshment station in 1652 led to the introduction of various varieties of 17th-century Dutch at the Cape. During the next 150 years, these Dutch speakers

came into contact with indigenous Khoekhoe, with slaves imported from Asia (India, Indonesia, Sri Lanka), East Africa, and Madagascar, and also, more sporadically, with French- and German-speaking Europeans. Written records reveal that a distinctive local variety of Dutch – so-called Kaaps Hollands (Cape Dutch), which was also variously described at the time as *geradbraakte/gebroke/onbeskaafde Hollands* ('mutilated/broken/uncivilized Hollandic'), *verkeerde Nederlands* ('incorrect Dutch') and *kom-buistaal* ('kitchen language') – already existed by the mid-18th century. There are three main positions on how this extraterritorial variety became a distinct, structurally simplified and reorganized language: the superstratist, variationist/interlectalist, and creolist positions. On the superstratist view, Afrikaans is essentially the product of the normal linguistic evolution that typically occurs in the absence of strong normative pressures, with the influence of Khoekhoe and the slave languages (i.e., Malay and Creole Portuguese) being confined to the lexical domain (see below). The variationist/interlectalist position similarly downplays the role of the non-Germanic languages interfacing with Dutch at the Cape, identifying dialect-leveling/convergence as the impetus behind the emergence of a new Dutch-based language. By contrast, the creolist view analyses Afrikaans as a semicreole, the product of interaction between the 'creolizing' and 'decreolizing' influences of the matrilectal Cape Dutch(es) and the Dutch-based pidgin(s) spoken respectively by the Cape's European and non-European populations. Exactly when Afrikaans was 'born' is also disputed, but official recognition of its distinctness came in 1925 when it was finally standardized following two *Taalbewegings* ('language movements') and recognized, alongside English, as one of South Africa's two official languages. The Bible was translated into Afrikaans in 1933 and a rich literary and cultural heritage accrued during the 20th century, with two major annual arts festivals now being dedicated solely to Afrikaans (the Klein Karoo Kunstefees/'Little Karoo Arts Festival' and Aardklop/'Earth-beat'). Because of its unfortunate association with the apartheid policy pursued between 1948 and 1994, there are, however, concerns about Afrikaans's future in post-apartheid South Africa and there has, in recent years, been a move to promote it as the only South African language which is both European *and* African.

Varieties of Afrikaans

The three basic varieties of Afrikaans traditionally identified are Kaapse Afrikaans (Cape Afrikaans) spoken in the western Cape, Oranjerivier-Afrikaans (Orange River Afrikaans) spoken in the northwestern

Cape, and Oosgrens-Afrikaans (Eastern Cape Afrikaans), the variety that provided the basis for standard Afrikaans, spoken in the rest of the country (see Figure 1). Kaapse and Oranjerivier Afrikaans are both spoken by people of color, the former reflecting particularly strong Malay and English influences, and the latter, that of Khoekhoe. Various subvarieties are discernible within these regional boundaries, one example being the Arabic-influenced Afrikaans spoken by Cape Muslims. Additionally, Afrikaans also forms the basis of a number of special group languages. Of these, Bantu-influenced Flaaitaal ('Fly-language'), a township argot spoken mostly by black migratory workers in urban areas, represents the best-studied case. During the apartheid era, normative pressures promoting suiwer Afrikaans ('pure Afrikaans') were strong and often directed against Anglicisms. Socio-political changes and attempts to promote Afrikaans as more 'inclusive' have, however, led to a more relaxed attitude in many contexts, with many younger speakers frequently speaking and writing Afrikaans, which is lexically heavily influenced by South Africa's other languages, particularly English. In its turn, Afrikaans has also left its mark on the other languages spoken in South Africa, with South African English featuring lexical items such as *braai* ('barbecue'), *veld* ('bush'), and *stoep* ('verandah'); Xhosa with *ispeki* (> *spek* = 'bacon'), *isitulu* (> *stoel* = 'chair'), and *ibhulukhwe* (> *broek* = 'trousers'); and Sotho, with *potloto* (> *potlood* = 'pencil'), *kerese* (> *kers* = 'candle'), and *sekotelopulugu* (> *skottelploeg* = 'disc-plough').

Formal Features

Many aspects of Afrikaans's formal structure represent simplifications of their Dutch counterparts, but the language also features a number of structural innovations. Phonologically, striking differences between Afrikaans and Dutch are that Afrikaans features:

- apocope of /t/ after voiceless consonants – cf. Afrikaans *lig* ('light') and *nag* ('night') versus Dutch *licht* and *nacht*
- syncope of intervocalic /d/ and /g/ – cf. Afrikaans *skouer* ('shoulder') and *spieël* ('mirror') versus Dutch *schouder* and *spiegel*
- fricative devoicing – cf. Afrikaans *suid* ('south') versus Dutch *zuid*
- diphthongization of long vowels – cf. Afrikaans [bruət] versus Dutch [bro:t] for *brood* ('bread').

There are also consistent orthographic differences, with Dutch *ij* and *sch* being rendered in Afrikaans as *y* and *sk*, respectively.

Morphologically, Afrikaans is characterized by extreme deflection: it lacks both Dutch's gender system



Figure 1 Map of South Africa showing the nine provinces created in 1994 and the areas in which the three main regional varieties of Afrikaans are spoken. Key: dark grey, Cape Afrikaans; light grey, Orange River Afrikaans; mid grey, Eastern Frontier Afrikaans.

and its system of verbal inflection, pronouns being the only nominals exhibiting distinct forms, although fewer than in Dutch (cf. Afrikaans *ons*, which corresponds to both Dutch *wij* – ‘we’ and *ons* – ‘us’), and all lexical verbs taking the same form, regardless of their person, number, and finiteness specifications. Afrikaans also differs from Dutch in employing reduplication – cf. *gou-gou* (‘quick-quick’), *stuk-stuk* (‘piece-piece,’ i.e., bit by bit), and *lag-lag* (‘laugh-laugh,’ i.e., easily).

Afrikaans’s retention of West Germanic’s distinctive word-order asymmetry (main clauses being verb-second/V2 and embedded clauses, verb-final) distinguishes it from Dutch-based creoles, which are exceptionlessly SVO and undermines extreme creolist accounts of its origins. Among the syntactic peculiarities that distinguish Afrikaans from Dutch are:

- its negative concord system – cf. Afrikaans *Ons lees nie hierdie boeke nie* (‘Us read not here – the books NEGATIVE’) and Dutch *Wij lezen niet deze boeken* (‘We read not these books’)
- verbal hendiads – cf. Afrikaans *Ek sit en skryf* (‘I sit and write’) versus Dutch *Ik zit te schrijven* (‘I sit to write,’ i.e., I sit writing)
- use of *vir* with personal objects – cf. *Ek sien vir jou* (‘I see for you’) versus Dutch *Ik zien je* (‘I see you’)

- *dat*-dropping in subordinate clauses – cf. *Hy weet ek is moeg* (‘He knows I am tired’), which alternates with *Hy weet dat ek moeg is* (‘He knows that I tired am’), whereas standard Dutch permits only the latter
- retention of main-clause ordering in subordinate interrogatives – cf. *Hy wonder wat lees ek* (‘He wonders what read I’) versus *Hy wonder wat ek lees* (‘He wonders what I read’), which is the only permissible structure in Dutch.

Lexically, Afrikaans differs substantially from Dutch in featuring borrowings from Khoekhoe, Malay, and Creole Portuguese (see ‘Lexical Borrowing’ section), and also, as a consequence of the ‘suiwer Afrikaans’ policy, in respect of many neologisms, which were created to avoid adopting an English expression – cf. *skemerkelkie*, *rekenaar*, and *trefferboek* or *blitsverkooper* whereas Dutch uses *cocktail*, *computer*, and *bestseller*, respectively.

The Taalmonument

Afrikaans is unique in being the only language with its own monument (see Figure 2). The *Taalmonument* (‘language-monument’) in Paarl was erected to celebrate the 100-year anniversary of the 1875 *Eerste Taalbeweging* (‘First Language-movement’) at which

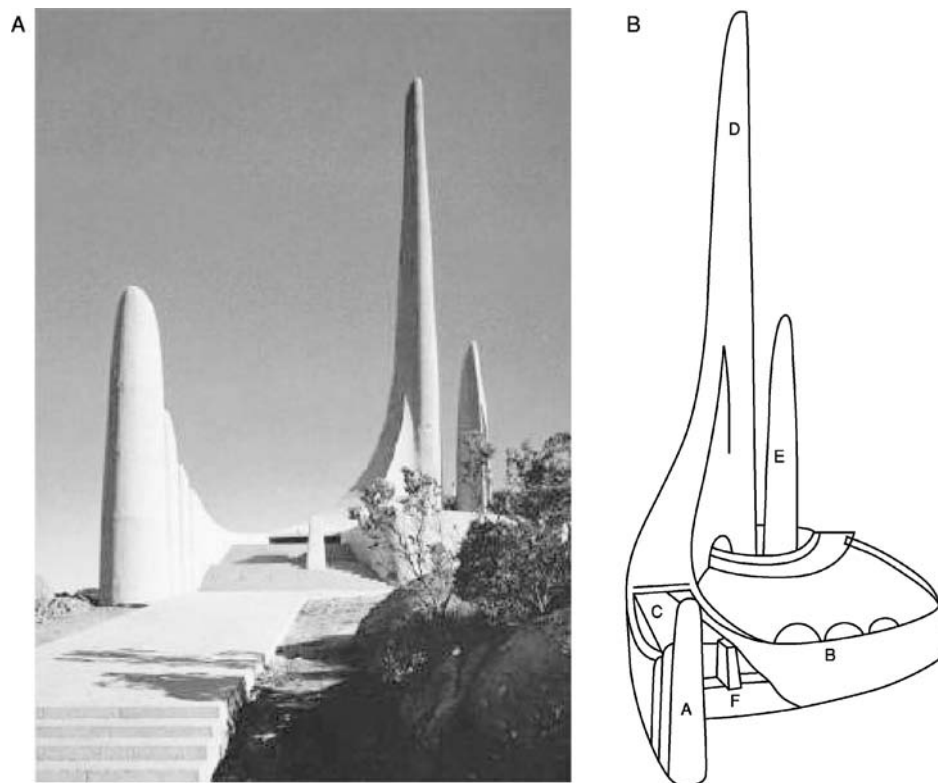


Figure 2 (A) The Afrikaans Language Monument (*Taalmonument*) in Paarl, South Africa. Reprinted by kind permission of the Afrikaans Language Museum, Paarl. (B) Diagrammatic representation of the structure of the Afrikaans Language Monument. A, The Enlightened West; B, Magical Africa; C, the bridge between the two; D, Afrikaans; E, The Republic of South Africa; F, Malay. Adapted from *Die Afrikaanse Taalmonument*, the official brochure of the Afrikaans Language Museum, Paarl.

the first concerted calls for the elevation of Afrikaans to the status of written language were made. The monument was inspired by the writings of two prominent Afrikaans writers, C. J. Langenhoven (1873–1832) and N. P. van Wyk Louw (1906–1970). Langenhoven visualized the growth potential of Afrikaans as a hyperbolic curve, whereas van Wyk Louw conceived of Afrikaans as “the language that links Western Europe and Africa ... form[ing] a bridge between the enlightened west and magical Africa” (1961, ‘Laat ons nie roem’/‘Let us not extoll’ in *Vernuwing in die Prosa/Renewal in prose*. Cape Town: Human and Rousseau). The monument symbolizes these ideas as follows:

- it features two curves (A and B) representing the influences of Europe and Africa respectively
- A, which starts as a colonnade, flows into the main column symbolizing Afrikaans (D), signifying the direct manner in which Afrikaans grew out of Dutch
- B, which features three semispherical mounds symbolizing the indigenous languages and cultures of South Africa, also flows into the main column via a lesser curve

- at the base of the column, A and B form a bridge (C) symbolizing the confluence of linguistic and cultural influences from Europe and Africa
- a low wall (F) located between A and B symbolizes the contribution of Malay
- column E represents the Republic of South Africa, the political entity established in 1961, within which Afrikaans was well established as one of two official languages.

Afrikaans was Written in Arabic

By the mid-19th century, Afrikaans was being used by the Cape Muslim community in the exercise of their religion and some of the *imams* were beginning to translate holy texts into Afrikaans using Arabic script. The first of these *ajami* (Arabic–Afrikaans) manuscripts, the *Hidāyat al-Islām* (‘Instruction in Islam’), is said to have been prepared in 1845 but is no longer extant. The first *ajami* text to be published, the *Bayānu ddīn* (‘Exposition of the religion’), was written by Abu Bakr in 1869 and published in Constantinople in 1877. Seventy-four texts, written between 1856 and 1957, survive today.

Lexical Borrowings

Afrikaans has drawn on the lexical resources of a wide variety of languages with which it has been in contact during the course of its history. Here are some examples of the range and nature of this borrowing:

- From Khoekhoe: animal names such as *geitjie* ('lizard'), *kwagga* (a zebra-like creature), and *gogga* ('insect'); plant names like *dagga* ('cannabis'); place names such as *Karoo* and *Knysna*; and also miscellaneous items such as *kierie* ('walking-stick'), *abba* ('carry') and *kamma* ('quasi/make-believe')
- From Malay: *baie* ('very/much'), *baadjie* ('jacket'), *baklei* ('fight'), *piesing* ('banana'), *rottang* ('cane'), *blatjang* ('chutney')
- From languages spoken on the Indian subcontinent: *koejawel* ('guava'), *katel* ('bed')
- From Creole Portuguese: *mielie* ('corn/maize'), *kraal* ('pen/corral'), *tronk* ('jail')
- From Bantu languages spoken in South Africa: *malie* ('money'), *aikôna* ('no'), *hokaai* ('stop'), *babelas* ('hangover').

Bibliography

den Besten H (1978). 'Cases of possible syntactic interference in the development of Afrikaans.' In Muysken P (ed.) *Amsterdam Creole Studies II*. Amsterdam: Instituut voor Algemene Taalwetenschap. 5–56.

den Besten H (1989). 'From Khoekhoe foreigntalk via Hottentot Dutch to Afrikaans: the creation of a novel grammar.' In Pütz M & Dirven R (eds.) *Wheels within wheels: Papers of the Duisburg Symposium on Pidgin and Creole Linguistics*. Frankfurt: Peter Lang. 207–249.

den Besten H (2002). 'Khoekhoe syntax and its implications for L2 acquisition of Dutch and Afrikaans.' *Journal of Germanic Linguistics* 14(1), 3–56.

den Besten H, Luijckx C & Roberge P (2003). 'Reduplication in Afrikaans.' In Kouwenberg S (ed.) *Twice as meaningful. Reduplication in pidgins, creoles and other contact languages*. London: Battlebridge. 271–287.

Botha R (1988). *Form and meaning in word formation: a study of Afrikaans reduplication*. Cambridge: Cambridge University Press.

Branford W & Claughton J (2002). 'Mutual lexical borrowings among some languages of southern Africa: Xhosa, Afrikaans and English.' In Mesthrie R (ed.). 199–215.

Buccini A (1996). 'New Netherland Dutch, Cape Dutch, Afrikaans.' *Taal en Tongval* 9, 35–51.

Davids A (1991). *The Afrikaans of the Cape Muslims from 1815 to 1915*. Master's thesis, Durban: University of Natal.

Deumert A (2002). 'Standardization and social networks: The emergence and diffusion of Standard Afrikaans.' In Linn A & McLelland N (eds.) *Standardization. Studies*

from the Germanic Languages. Amsterdam: John Benjamins. 1–25.

Deumert A (2004). *Language standardization and language change: the dynamics of Cape Dutch*. Amsterdam: John Benjamins.

Donaldson B (1991). *The influence of English on Afrikaans: a case study of linguistic change in a language contact situation*. Pretoria: Academica.

Donaldson B (1993). *A grammar of Afrikaans*. New York: Mouton de Gruyter.

Donaldson B (1994). 'Afrikaans.' In König E & van der Auwera J (eds.) *The Germanic languages*. London: Routledge. 478–504.

Holm J (2001). 'The semicreole identity of Afrikaans lects: parallel cases of partial restructuring.' *Journal of Germanic Languages* 13(4), 353–379.

Kotzé E (2001). 'Adjectival inflection in Afrikaans diachronics: an argument against the validity of creolization checklists.' *Journal of Germanic Languages* 13(4), 380–399.

Lass R & Wright S (1986). 'Endogeny vs. contact: "Afrikaans influence" on South African English.' *English World-Wide* 7, 201–224.

Makhudu K (2002). 'An introduction to Flaaitaal (or Tsotsitaal).' In Mesthrie R (ed.). 398–406.

McCormick K (2002a). *Language in Cape Town's District Six*. Oxford: Oxford University Press.

McCormick K (2002b). 'Code-switching, mixing and convergence in Cape Town.' In Mesthrie R (ed.). 216–234.

Mesthrie R (ed.) (2002). *Language in South Africa*. Cambridge: Cambridge University Press.

Ponelis F (1993). *The development of Afrikaans*. Frankfurt: Peter Lang.

Raidt E (1983). *Einführung in Geschichte und Struktur des Afrikaans*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Reagan T (2002). 'Language planning and language policy: past, present and future.' In Mesthrie R (ed.). 419–433.

Rensburg C van (1989). 'Orange River Afrikaans: a stage in the pidgin/creole cycle.' In Pütz M & Dirven R (eds.) *Wheels within wheels: Papers of the Duisburg Symposium on Pidgin and Creole Linguistics*. Frankfurt: Peter Lang. 135–151.

Roberge P (1994). *The formation of Afrikaans*. Stellenbosch: University Printers.

Roberge P (2000). 'Etymological opacity, hybridization, and the Afrikaans brace negation.' *American Journal of Germanic Linguistics and Literatures* 12(1), 101–176.

Roberge P (2002a). 'Afrikaans: considering origins.' In Mesthrie R (ed.). 79–103.

Roberge P (2002b). 'Convergence and the formation of Afrikaans.' *Journal of Germanic Linguistics* 14(1), 57–94.

Stone G (2002). 'The lexicon and sociolinguistic codes of the working-class Afrikaans-speaking Cape Peninsula coloured community.' In Mesthrie R (ed.). 381–397.

Thomason S & Kaufman T (1991). *Language contact, creolization, and genetic linguistics*. Berkeley, Los Angeles & Oxford: University of California Press. 251–256.

Valkhoff M (1966). *Studies in Portuguese and Creole: with special reference to South Africa*. Johannesburg: Witwatersrand University Press.

Afroasiatic Languages

J Crass, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Afroasiatic languages are spoken by more than 250 million people living in northern Africa, the Horn of Africa, and in South West Asia. The Afroasiatic language phylum (or superfamily) contains more than 200 languages, even 372 according to Grimes (2000). In addition, a number of languages are documented only literally. With the exception of the extinct Sumerian, Afroasiatic has the longest documented history of any language phyla in the world: Egyptian was recorded as early as 3200 B.C., while the documentation of Semitic languages goes back to 2500 B.C. The name Afroasiatic was established by Greenberg (1952), replacing the inappropriate term Hamito-Semitic (or rarely Semito-Hamitic) that is still used by a few scholars. Other terms with little acceptance are Afrasian, Erythraic, and Lisramic.

Classification and Geographical Origin

The Afroasiatic languages are divided into six branches, namely Berber, Chadic, Cushitic, Egyptian, Omotic, and Semitic. Whereas Egyptian (Arabic, Egyptian Spoken) is a single language with four stages (Old-, Middle-, and New-Egyptian and Coptic), the other five branches are families. Chadic encompasses the largest number of languages – namely 195 according to Grimes (2000) or approximately 140 according to Newman (1992) – followed by Semitic (74), Cushitic (47), Omotic (28), and Berber (26), the latter four numbers as stated by Grimes (2000). These six branches are considered ‘sister families,’ i.e., they are equal, flat, and parallel. However, there are attempts to connect these branches to larger units. Semitic and Berber are relatively closely related, and both are somehow connected to Cushitic (Zaborski, 1997). Bender (1997) calls this group of branches macro-Cushitic and speculates on its connection with Indo-European.

According to Diakonoff (1988) and Bender (1997), the original homeland of the speakers of Afroasiatic languages was in the southeast of today’s Saharan desert, while Militarev and Shnirelman (1984) believe it was in Asia. The former scenario seems likely because – except for Semitic – all families of the Afroasiatic phylum are spoken exclusively in Africa. The latter scenario is also possible, however, because parts of the lexis are shared by the Afroasiatic

languages, the Sumerian language, and the Caucasian languages (Hayward, 2000: 95).

History of the Investigation of Afroasiatic Languages

In the Middle Ages, the genetic relationship between the Semitic languages Arabic (Standard Arabic) and Hebrew was discovered only after the study of Afroasiatic languages had already begun. Likewise, only after Egyptian was deciphered in the 19th century did the affinity of Egyptian to Semitic become apparent. A short time later, Berber and Cushitic were recognized as belonging to this phylum. The Chadic languages as a whole were classified as Afroasiatic languages by Greenberg in the 1950s. The sixth branch, Omotic, was regarded as a branch of Cushitic until the end of the 1960s, and while some scholars still consider this to be true (Lamberti, 1991; Zaborski, 1986, 1997), most believe that Omotic is an independent branch of Afroasiatic (Fleming, 1969). A few scholars even regard it as the first family that split off from Proto-Afroasiatic, the reconstructed ancestor of all Afroasiatic languages (Fleming, 1983; Ehret, 1995).

Finally, it should be mentioned that Hetzron (1980) sees Beja (Bedawi) – generally regarded as the only representative of North Cushitic – as another family of Afroasiatic, but Zaborski (1984) does not agree with this view.

For a long time, the structure and features of Semitic determined which languages belonged to the Afroasiatic language phylum. Most likely this was because Arabic and Hebrew were the first languages European scholars knew. Also, for a significant period of time, racial, even racist prejudices dominated classification suggestions of the Afroasiatic languages. In the mid–19th century, the idea of a language family, of which Semitic is one branch, was born. The term Hamitic, derived from the name Ham, the second son of Noah, was created in opposition to Sem, the name of the first son of Noah, who was the eponym of the Semitic languages. All Afroasiatic languages related to Semitic, but considered to be non-Semitic, were classified as Hamitic, the second branch of ‘Hamito-Semitic.’ These criteria were a mixture of linguistic (genetic and typological), physical anthropological, and partly geographical features.

Lepsius (1863), the first important exponent of this theory, classified the Hamitic branch into four groups, namely (1) Egyptian; (2) Ethiopic (Ge’ez), i.e., mostly Cushitic languages spoken in the Horn of Africa; (3) Libyan, i.e., Berber and the Chadic language

Hausa; and (4) Hottentottan (Nama), i.e., languages of the Khoisan phylum of southern and southwestern Africa. In 1880 he included even Maasai – a language of the Nilosaharan phylum – in the Hamitic branch. Lepsius's main criterion for his classification was grammatical gender. African languages possessing the masculine vs. feminine gender distinction were classified Hamitic, while African languages without gender distinction were called 'Negersprachen,' i.e., 'languages of the negroes.'

The most famous exponent of the Hamitic theory was Meinhof (1912), who tried to work out the features of the Hamitic languages by considering genetic, typological, and physical anthropological features. Meinhof was of the opinion that one must distinguish more 'primitive' from more 'highly developed' languages, a criterion that he believed correlated with the mental abilities of the speakers of the respective languages. In the tradition of Schleicher, he believed that inflecting languages reflect the highest level of linguistic evolution. This typological feature of the Hamitic languages derived from a race called 'Hamites' who had white skin, curled hair, and other physical anthropological features considered prototypical of the old Egyptian and Ethiopide types.

Besides grammatical gender, ablaut and other typological features of the Indo-European and Semitic languages were the main linguistic criteria Meinhof took into consideration. He classified as Hamitic not only Afroasiatic languages (except Semitic) but also languages like Ful (Fulfulde, Adamawa) (an Atlantic language of the Niger Congo phylum), Maasai, and other Nilotic languages of the Nilosaharan phylum and languages of the Khoisan phylum, earlier excluded by others from the Afroasiatic languages.

The first opponents of the Hamitic theory were Beke (1845) and Lottner (1860–61), later followed by Erman (1911) and Cohen (1933) who considered – as did the aforementioned scholars – the branches of this phylum to be 'sister families.' According to Sasse (1981: 135), the final breakthrough of this theory and the beginning of a new era in the study of Afroasiatic languages was marked by Cohen (1947). Greenberg (1952, 1955) finally provided evidence that a number of languages had to be excluded from the Afroasiatic language phylum, and he created the Chadic family by unifying the former 'chadohamitic' language Hausa with the rest of the Chadic languages that until then had been classified as non-Afroasiatic languages.

Shared Features

The genetic relationship among the six branches of Afroasiatic is shown best by some shared morphological features (cf. Hayward, 2000: 86ff; Sasse, 1981:

138ff). These are case marking, plural formation on nouns, gender marking, pronouns, verb inflection, and verb derivation.

The basic nominal form of Proto-Afroasiatic, functioning as the direct object of a verb, is termed 'absolute,' marked by the suffix **-a*. In Cushitic and – as Sasse (1984) claims – in Semitic and Berber, its function is more widespread, so it can be treated as the functionally unmarked form. The nominative, marked by **-u*, is used for subject NPs. A similar morphology can be assumed for Egyptian and Omotic, the latter having a reconstructed accusative marking system (Hayward and Tsuge, 1998), i.e., the unmarked form is the nominative and not – as reconstructed for Semitic, Berber, and Cushitic – the absolute. Chadic, however, is not concerned here since it generally lacks case marking. Modern languages with a marked nominative case system occur mainly in central and southwestern Ethiopia and adjacent areas where this system of case marking is an areal feature found not only in several Cushitic and Omotic languages, but also in languages of the Nilosaharan phylum.

Complex plural formation of nouns is another characteristic of many Afroasiatic languages. A likely pattern of Afroasiatic plural formation is the "ablaut to *a*, usually in the last stem syllable of a noun ... [partly] accompanied by reduplication, and sometimes trigger[ing] dissimilation or assimilation of other stem vowels of the plural" (Hayward, 2000: 92; cf. Greenberg, 1955). Other reconstructed plural markers are a suffix containing a labial-velar glide and a suffix *-t*, the latter not easy to disentangle from the *-t* of the feminine gender marker. Such a gender marker is found, in all six branches of Afroasiatic. In addition gemination of consonants marking nominal and verbal plurality is widespread.

Two formally distinct sets of pronouns must be set up for Afroasiatic, the first for the absolute, the second for the nominative case. Due to the shift of a marked nominative to a marked accusative system, the absolute pronouns often were converted to nominative pronouns, e.g., in Berber and Chadic, so consequently, the subject pronouns of these languages just happen to look like object pronouns of other languages. Gender markers **n-* and **k-* for masculine and **t-* for feminine are often derived from demonstrative elements. These gender markers may be combined with the pronominal gender marker **-uu* for masculine and **-ii* for feminine and function as demonstrative pronouns, especially of the near deixis. This applies exactly to the Highland East Cushitic language K'abeena, in which the demonstrative pronouns have an additional morpheme *n* – probably a definite marker – that results in the forms *kuun* and *tiin*.

Subject agreement on the verb may be marked in two ways, either by a so-called prefix conjugation or by a suffix (or stative) conjugation. Some languages make use of both, e.g., most modern Semitic languages; others have only the suffix conjugation, e.g., Egyptian and many Cushitic languages. The reconstructed subject-agreement morphemes of the prefix conjugation are *ʾ- (1S), *t- (2S, 3Sf, 2P), *y- (3Sm), and *n- (1P). Suffixes differentiate number and partly gender.

Some morphemes used for verb derivation are found in many Afroasiatic languages, so most probably those are a feature of Proto-Afroasiatic. The transitivity/causativizing *s- ~ *-s and the intransitivity/passivizing *m- ~ *-m, *n-, and *t- ~ *-t belong to these morphemes.

Furthermore, hundreds of lexical items have been reconstructed for Proto-Afroasiatic by Ehret (1995) and Orel and Stolbova (1995) of which a small number “seem unlikely to be disputed” (Hayward, 2000: 94), e.g., *dim-/dam- ‘blood’, *tuf- ‘to spit’, *sum-/sim- ‘name’, *sin-/san- ‘nose’, *man-/min- ‘house’, and *nam-/nim- ‘man’.

The rich consonant inventory of Proto-Afroasiatic – Orel and Stolbova 1995: xvi reconstruct 32, Ehret 1995: 72, even 42 consonants – includes three obstruents, namely, a voiceless, a voiced, and a glottalized sound “not only for most places of articulation but also for certain other articulatory parameters, for example, among lateral obstruents, sibilants and labialised velars” (Hayward, 2000: 94). Furthermore, two pharyngeals, two glottals, and four uvulars are reconstructed by Orel and Stolbova (1995).

Typologically, there is a contrast between Berber, Egyptian, and Semitic on the one hand and Chadic, Cushitic, and Omotic on the other. According to Bennett (1998: 22), the first three languages “generally have (or can be reconstructed as having had) three underlying vowels, no tonal contrasts ... and typically triconsonantal roots that at least in the verbal system seem not to include vowels.” He writes that the latter three, however, are characterized by “relatively full vowel systems, tonal contrasts, and roots of varied length that normally include a vowel” (Bennett, 1998: 22). Concerning word order, Afroasiatic languages can be divided as follows: Berber, Chadic, and Semitic languages outside Ethiopia have VO word order, while Cushitic, Omotic, and Ethiop-Semitic languages generally have OV word order.

Finally, two hypotheses must be mentioned. Diakonoff (1965) is of the opinion that Proto-Afroasiatic was an ergative language, a hypothesis adopted by Bender (1997) and for Semitic by Waltisberg (2002). The second hypothesis concerns the possible substrate influence of Afroasiatic

languages on the Celtic languages (cf. Adams, 1975; Gensler, in press).

Bibliography

- Adams G B (1975). ‘Hamito-Semitic and the pre-Celtic substratum in Ireland and Britain.’ In Bynon J & Bynon T (eds.) *Hamito-Semitic: proceedings of a colloquium held by the Historical Section of the Linguistics Association (Great Britain) at the School of Oriental and African Studies, University of London, on the 18th, 19th and 20th of March 1970*. The Hague, Paris: Mouton. 233–247.
- Beke C T (1845). ‘On the languages and dialects of Abyssinia and the countries to the south.’ *Proceedings of the Philological Society* 2, 89–107.
- Bender L M (1997). ‘Upside-down Afrasian.’ *Afrikanistische Arbeitspapiere* 50, 19–34.
- Bennett P R (1998). *Comparative Semitic linguistics: a manual*. Winona Lake: Eisenbrauns.
- Bynon J (ed.) (1984). *Current progress in Afro-Asiatic linguistics: papers of the Third International Hamito-Semitic Congress*. Amsterdam, Philadelphia: Benjamins.
- Cohen M (1933). *Les Résultats acquis de la grammaire comparée chamito-sémitique*. Paris: Champion.
- Cohen M (1947). *Essai comparatif sur le vocabulaire et la phonétique du chamito-sémitique*. Paris: Champion.
- Diakonoff I M (1965). *Semito-Hamitic languages: an essay in classification*. Moscow: Nauka.
- Diakonoff I M (1988). *Afrasian languages*. Korolevna A A & Porkhomovsky V J (trans.). V. J. Moscow: Nauka.
- Ehret C (1995). *Reconstructing Proto-Afroasiatic (Proto-Afrasian): vowels, tone, consonants and vocabulary*. Berkeley: University of California Press.
- Erman A (1911). *Ägyptische grammatik* (3rd edn.). New York: Westermann.
- Fleming H (1969). ‘The classification of West Cushitic within Hamito-Semitic.’ In Butler J (ed.) *East African History*. New York: Praeger. 3–27.
- Fleming H (1983). ‘Chadic external relations.’ In Wolff E & Meyer-Bahlburg H (eds.) *Studies in Chadic and Afroasiatic linguistics*. Hamburg: Buske. 17–31.
- Gensler O (in press). *The Celtic-North African linguistic link: substrata and typological argumentation*. Oxford: Oxford University Press.
- Greenberg J H (1952). ‘The Afroasiatic (Hamito-Semitic) present.’ *Journal of the American Oriental Society* 72, 1–9.
- Greenberg J H (1955). *Studies in African linguistic classification*. New Haven: Compass.
- Grimes B (2000). *Ethnologue, vol. 2: maps and indexes*. Dallas: SIL International.
- Hayward R (2000). ‘Afroasiatic.’ In Heine B & Nurse D (eds.) *African languages: an introduction*. Cambridge: Cambridge University Press. 74–98.
- Hayward R & Tsuge Y (1998). ‘Concerning case in Omotic.’ *Afrika und Übersee* 81, 21–38.
- Hetzron R (1980). ‘The limits of Cushitic.’ *Sprache und Geschichte in Afrika* 2, 7–126.

- Lamberti M (1991). 'Cushitic and its classifications.' *Anthropos* 86, 552–561.
- Lepsius R (1863). *Standard alphabet for reducing unwritten languages and foreign graphic systems to a uniform orthography in European letters*. London: Williams and Norgate.
- Lepsius R (1880). *Nubische Grammatik. Mit einer Einleitung über die Völker und Sprachen Afrikas*. Berlin: Hertz.
- Lottner C (1860–61). 'On sister families of languages, specially those connected with the Semitic family.' *Transactions of the Philological Society* 20–27, 112–132.
- Meinhof C (1912). *Die Sprache der Hamiten*. Abhandlungen des Hamburgischen Kolonialinstitutes 9. Hamburg: Friederichsen.
- Militariyev A J & Shnirelman V A (1984). *K probleme lokalosatsii drevneishikh afraziitsev: opyt lingvoarkheologicheskoi rekonstruktsii*. (On the problem of location of the early Afrasians: an essay in linguo-archeological reconstruction.) *Lingvisticheskaya rekonstruktsiya i drevneyshaya istoriya vostoka*. Chast' 2. Moscow.
- Newman P (1992). 'Chadic languages.' In Bright W (ed.) *International encyclopedia of linguistics*. New York, Oxford: Oxford University Press. 251–254.
- Orel V E & Stolbova O V (1995). *Hamito-Semitic etymological dictionary: materials for a reconstruction*. Leiden: Brill.
- Sasse H-J (1981). 'Afroasiatisch.' In Heine B, Schadeberg T C & Wolff E (eds.) *Die Sprachen Afrikas*. Hamburg: Buske. 129–148.
- Sasse H-J (1984). 'Case in Cushitic, Semitic and Berber.' In Bynon James (ed.). 111–126.
- Waltisberg M (2002). 'Zur Ergativitätshypothese im Semitischen.' *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 52(1), 11–62.
- Zaborski A (1984). 'Remarks on the genetic classification and the relative chronology of the Cushitic languages.' In Bynon J (ed.). 125–138.
- Zaborski A (1986). 'Can Omotic be reclassified as West Cushitic?' In Goldenberg G (ed.) *Proceedings of the 6th International Conference of Ethiopian Studies, Tel Aviv 14th to 17th April, 1980*. Rotterdam: Balkema. 525–529.
- Zaborski A (1997). 'The position of Cushitic and Berber within Hamitosemitic Dialects.' In Bausi A & Tosco M (eds.) *Afroasiatica neapolitana: papers from the 8th Italian Meeting of Afroasiatic (Hamito-Semitic) Linguistics, Naples, January 25–26, 1996*. Napoli: Istituto Universitario Orientale.
- Zaborski A (1999). 'On the alleged ergativity in Hamito-Semitic/Afroasiatic languages.' In Brzezina M & Kurek H (eds.) *Collectanea linguistica. In honorem Casimiri Polanski*. Kraków: Księgarnia Akademicka. 309–317.

Ainu

M Shibatani, Rice University, Houston, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

Ainu is a near-extinct language that was once spoken widely in the northern part of the main Japanese island of Honshu as well as the Hokkaido island, in Sakhalin, and in the Kurile Islands. The current Ainu population, concentrated mainly in Hokkaido, is estimated to be around 24 000, but as a result of intermarriage between Ainu and Japanese, pure-blood Ainu are said to number less than 1% of that figure. Ainu is no longer used as a means of daily communication, and is remembered to a varying extent only by a handful of people of advanced age.

Ainu has not developed a writing system, but it is endowed with a rich tradition of oral literature. In addition to various kinds of songs, e.g., love songs and boating songs, Ainu oral literature contains both verse and prose. The verse forms, generally called *yukar* in Ainu, are recited epics that relate to the experiences of gods or to the experiences of love and war of heroes. The language of *yukar* differs

significantly from the spoken language; it is more conservative and has less dialectal variation as compared with the colloquial language. The two types of language show differences in both syntax and vocabulary, although there is a great deal of overlap. The most salient difference between the two is that the language of *yukar* tends to be more strongly polysynthetic than its colloquial counterpart. The language of *yukar* will be referred to as Classical Ainu, but the difference between this type of language and the colloquial form is more a difference in genre than in chronology.

In terms of genetic affiliation, Ainu is best considered as a language isolate. Although there have been suggestions that Ainu is related to such language families as Paleo-Asiatic, Ural-Altaic, Indo-European, and Malayo-Polynesia or to individual languages such as Gilyak and Eskimo, none of these suggestions has progressed beyond the level of speculation. Hypotheses relating Ainu to Japanese have also been entertained by many scholars, but other than the similarities due to lexical borrowing and typological characteristics rooted in the shared basic word order (Subject–Object–Verb), no strong evidence has been

uncovered to relate the two languages. Indeed, Ainu has a number of morphological characteristics that distinguish it from Japanese, e.g., extensive use of personal affixes and a polysynthetic character as well as absence of verbal inflections.

Ainu has a rather simple phonological system, with five vowel phonemes (/i, e, a, o, u/) and 12 consonantal phonemes (/p, w, m, t, s, c, y, n, r, k, ʔ, h/). Syllable-initial vowels are preceded by a glottal stop, e.g., *aynu* [ʔajnu] ‘person,’ and this fact makes Ainu syllables conform to one of the following types: CV, CVC (for Hokkaido Ainu) or CV, CVV (long vowel), CVC (for Sakhalin Ainu).

According to the pitch accent system of the language, Ainu syllables are pronounced with either high or low pitch. In words consisting of stems and affixes, the stems have high pitch, e.g., *nú-pa* ‘to hear-pl.OBJ.’ In other two- and three-syllable words, high pitch falls on the first syllable if it is a heavy syllable, i.e., a diphthong or a closed syllable, e.g., *áynu* ‘person.’ In all other words, high pitch occurs in the second syllable, e.g., *kirá* ‘to flee.’

Among the small number of phonological processes, the most notable are assimilatory and dissimilatory processes of the following type: *akor nispa* → *akon nispa* ‘our chief,’ *pon-pe* → *pompe* ‘small thing,’ (assimilation); *kukor rusuy* → *kukon rusuy* ‘want to have’ (dissimilation).

Both nominal and verbal morphologies are characterized by extensive use of affixes. In nominal morphology perhaps the most notable are deverbal nominal suffixes that derive nominal expressions from verbs. The suffix *-p(e)* derives a noun that denotes a person or things characterized by the meaning of the original verb, e.g., *pirka* ‘good’ → *pirka-p* ‘good thing,’ *wen* ‘bad’ → *wen-pe* ‘poor man.’

Two other noun-forming derivational affixes are the suffixes *-i* and *-ike*. The former yields nouns having the meaning ‘X-place’ or ‘X-time,’ and the latter produces nouns with the meaning ‘thing’ or ‘person,’ e.g., *esan* ‘go out there’ → *esan-i* ‘place that is protruded, i.e., peninsula,’ *poro* ‘big’ → *poro-ike* ‘bigness, big thing/person.’

One notable feature of these suffixes with theoretical significance is that they, especially *-p(e)* and *-i*, also attach to phrases and clauses, functioning as both lexical and phrasal nominalizing suffixes, e.g., *a-koyki rok-pe* (1sg-strike PERF-SUF) ‘the one I have fought,’ *a-yanene-p ya-kotan-oro esina-p* (1sg-dislike-SUF REFL-village-from hide-SUF) ‘what I dislike is hiding one’s village (from which one came).’

Possession is expressed by the use of personal affixes that, when attached to verbs, index the subject

of transitive clauses, e.g., *a-maci* (1sg-wife), *e-maci* (2sg-wife) ‘young wife,’ *maci* ‘his wife.’

In both Classical and colloquial Ainu, intransitive and transitive verbs each have distinct sets of personal affixes indicating person and number of the subject and object, e.g., Classical Ainu intransitive affixes: *itak-an* (speak-1sg) ‘I speak,’ *e-itak* (2sg-speak) ‘you (sg) speak,’ *itak* ‘he/she speaks’; Classical Ainu transitive affixes: *a-kor* (1sg-have) ‘I have,’ *e-kor* (2sg-have) ‘you (sg) have,’ *kor* ‘he/she has.’ These subject-indexing affixes combine with object-indexing affixes, yielding forms such as *a-e-kore* (1sg-2sg-give) ‘I give you,’ *e-i-kore* (2sg-1sg-give) ‘you give me.’

Ainu verbs – Ainu makes no distinction between verbs and adjectives – also index the plurality of the subject and object. The plural verb forms typically co-occur with a plural subject when the verb is intransitive and with a plural object when it is transitive. However, Ainu also shows cases of plural verbs co-occurring with plural transitive subjects. Plural verbs are of either suppletive type (*arpa* ‘go,’ *paye* ‘go.pl’) or productive-suffixed type (*kor* ‘have (sg)’; *kor-pa* ‘have (pl)’); e.g., *An-an* (be-1sg) ‘I was (there)’; *Oka-an* (be.pl-1pl) ‘We were (there)’; *Icen poronno kor-pa* (money lot have-pl) ‘They had a lot of money’ (Ishikari dialect); *Sisam sokor goza sinep hok-pa wa arki* (Japanese from mat one buy-pl and come.pl) ‘They bought one mat from a Japanese and came’ (Ishikari dialect).

Plural verb forms are also used as honorifics, e.g., *Kane rakko a-res-pa kamuy ronnu* (golden otter 1pl-raise-pl god kill.pl) ‘Our honorable god, whom we have raised, killed the golden sea otter’.

The most notable feature of Ainu verbal morphology is incorporation of various elements – the feature that contributes to the polysynthetic character of Ainu, especially Classical Ainu. Nouns corresponding to intransitive subjects and those corresponding to transitive objects are incorporated, though many instances of the former type appear to be frozen expressions, e.g., *Sir-pirka* (weather-good) ‘It’s fine.’ Typical noun incorporation is of the following type, where incorporation of a noun corresponding to an object results in an intransitive expression with concomitant change in the personal affix: *Cise ci-kar* (house 1pl-make) ‘We make a house’; *Cise-kar-as* (house-make-1pl) ‘We make a house’ (Ishikari dialect).

In addition, Ainu verbs incorporate adverbs, e.g., *Toyko a-kikkik* (thoroughly 1sg-beat) ‘I beat (him) up thoroughly’; *A-toyko-kikkik* (1sg-thoroughly-beat). While no more than one noun can be incorporated into the verb at a time, a noun and an adverb can be

incorporated into one verb base at the same time, e.g., *Pinne kamuy kiraw-rik-kur-roski* (male god horn-high-EXPL-raise) 'The male (dragon) god raised the horns high.'

Moreover, Ainu verbal morphology permits applicative extension, thereby exhibiting the following paraphrases between postpositional expressions and the corresponding applicative expressions: *Porō cise ta horari* (big house at live) 'He lives in a big house': *Porō cise e-horari* (big house APPL-live) 'He lives in a big house'; *kaya ari terke* (sail with run) 'run by a sail': *kaya e-terke* (sail APPL-run) 'run by a sail.'

A combination of noun incorporation and applicative extension yields an expression such as *Nea cep a-pone-ko-kuykuy* (that fish 1sg-bone-APPL-bite) 'I bit that fish with its bone.'

Ainu syntax is consistently head-final, thereby exhibiting word order patterns similar to those observed in other head-final languages such as Japanese and Korean. Thus, the basic word order is SOV: *Kamuy aynu rayke* (bear person kill) 'The bear killed the man.' Postpositions are used rather than prepositions: *cise ta* (home at) 'at home,' and modifiers precede the heads they modify: *pirka kewtum* (good heard) 'good heart,' [*beko respa*] *sisam* ([cow raise] Japanese) 'a Japanese who raises cows,' *sapo ninkarihi* (sister earrings) 'sister's earrings,' *toan seta* (that dog) 'that dog,' *sine aynu* (one person) 'one

person,' *turasno paye* (quickly go) 'go quickly,' *a-e rusuy* (1sg-eat want) 'want to eat,' *menoko kasuno okirasunu* (woman than strong) 'stronger than a woman.'

Subordinating conjunctions occur after subordinate clauses, which come before main clauses, e.g., *E-eh kusu anekiroro-an* (2sg-came because happy-1sg) 'Because you came, I am happy' (Sakhalin dialect).

Auxiliary verbs are not generally marked by personal affixes, which are attached to the main verbs. And finally, question sentences are marked by the final particle *ya*, or are simply indicated by rising intonation alone. Like many other head-final languages, interrogative pronouns need not move to sentence-initial position. The following final example illustrates the use of auxiliary verbs and interrogative sentence pattern: *Eani hemanta e-e rusuy ya* (you what 2sg-eat want Q) 'What do you want to eat?'

Bibliography

- Batchelor J ([1038] 1981). *An Ainu-English-Japanese dictionary* (4th edn.). Tokyo: Iwanami.
- Refsing K (1986). *The Ainu language – the morphology and syntax of the Shizunai dialect*. Copenhagen: Aarhus University Press.
- Shibatani M (1990). *The languages of Japan*. Cambridge: Cambridge University Press.
- Tamura S (2000). *The Ainu language*. Tokyo: Sanseido.

Akan

M E Kropp Dakubu, University of Ghana, Legon, Accra, Ghana

© 2006 Elsevier Ltd. All rights reserved.

The Akan language is spoken throughout the central portion of Ghana. It is the most widely spoken member of a family of about 20 languages known as Tano or Volta-Comoe spoken in Ghana and the eastern Ivory Coast. Formerly the entire group was referred to as Akan. These languages belong to the Niger-Congo family. Within Niger-Congo they are part of the Kwa grouping.

Dialects and Their Distribution

The name 'Akan' is not generally used by speakers of the language, who refer to their language as Fante, Twi, or Brong. These Akan speech forms constitute a dialect continuum running from north to south in Ghana. 'Fante' refers to the dialects spoken in

those regions that reach the sea, in the Central Region and parts of the Western Region of Ghana. 'Twi' is the most general term, referring to a wide range of dialects, of which the best known are Akuapem, the main tongue of the Eastern Region, and Asante, the dialect of the Ashanti Region. Others are Akyem and Kwahu. In genetic terms, Akuapem is more closely related to Fante than to the other dialects, but all of these dialects are mutually intelligible. The Brong dialect group of the Brong-Ahafo Region to the north of Ashanti is mutually intelligible with Asante Twi, but there is less mutual intelligibility with the dialects spoken farthest south.

History and Development

Lists of several hundred words in Fante were published in Europe during the 17th and 18th centuries, but the language became a written language

with a printed literature in the first half of the 19th century. The first written form was based on the Akuapem dialect, and was the work of members of the Basel Mission, which became established in the Eastern Region in the 1830s. The major names connected with this work are H. N. Riis, who published the first grammar in German in 1853 and in English in 1854, and Johann Gottlieb Christaller, whose grammar and dictionary appeared in 1875 and 1881, respectively. His collection of 3,600 Akan proverbs appeared in 1879. Christaller's work was important not only for Akan but for West African linguistics generally, because he analyzed the characteristic vowel harmony system and the tone system (see later), and their significance for the grammar.

The Akuapem-based orthography was used in schools of the Basel Mission, and later throughout the Twi-speaking areas until an Asante orthography was established in the 1950s. Since then, three orthographies, Fante, Asante, and Akuapem, have been used in the schools. A Unified Akan Orthography was developed in 1978 and published, but has not been put into practice by publishers or teachers. Nevertheless, more works have been published in Akan than in any other Ghanaian language, more than half of them in the Akuapem orthography.

Sociolinguistic Situation

As mother tongue of about 43% of the population of Ghana (7 550 405 out of about 18 million) and spoken as a second language by many more, Akan is indisputably the most commonly spoken Ghanaian language. Asante, with 2 578 829 speakers, is the largest dialect, Fante coming second with 1 723 573 speakers (figures are based on the report of the 2000 Census). Exactly how many speak Akan as a second language is not known, but there are very few places in Ghana where a speaker cannot be found. The Asante dialect seems to be the most widely known, and is expanding. Although Accra, the capital of Ghana, historically is not an Akan town, there are strong indications that today Akan is more widely spoken there than any other Ghanaian language.

From the 17th century until British conquest in the 20th century, Akan was the language of expanding kingdoms, of which the Ashanti became the largest and most famous. The resulting impact on the other languages of Ghana was considerable, especially in the south. Virtually all southern Ghanaian languages have borrowed Akan words related to war, government/state, the arts (especially music), and personal names and appellations. Akan is the

source of several English words and proper names, especially in the Caribbean. The most well-known English word of Akan origin is probably the name of the Jamaican folktale character, Anancy, from Akan *ananse* 'spider'. Another is okra, from Akan *ɛ-koro-ma*.

Akan is the language most used after English in the electronic public media, and in some areas is used more than English. This is most noticeable on the FM radio stations distributed throughout Akan-speaking regions and in Accra. It is fairly often heard on television and is very commonly used in both television and radio advertising. However, there is little if any print journalism in Akan, although there has been more in the past.

Akan is a school subject in Akan-speaking regions, in many Accra schools, and in teacher training colleges. It can be studied to degree level at the University of Ghana and the University of Cape Coast, and is an area of specialization at the University College of Education at Winneba.

Aspects of the Ethnography of Speaking

Formal speech is very important in Akan culture. Every chief or king has an *ɔkyeame*, or spokesman, whose function is to speak for the chief on all formal occasions. This man is highly regarded as a master of the language. Elegant speech, especially that used at court, is profuse and indirect. Mastery of proverbs and their appropriate use are important aspects of this style.

Major Linguistic Features

The Sounds of Akan

This section is based mainly on Dolphyne's (1988) *The Akan (Twi-Fante) language*, which should be consulted for more detail.

Consonants The Akan consonants p, b, t, d, k, g, m, n, f, s, h, w, l, r, and y are usually pronounced much as they are in English, although n is pronounced [ŋ] in some contexts, e.g., in *nkwana* 'soup'. The spellings ky, gy, and hy, however, are pronounced similarly to English ch, j, and sh, respectively. Akan also has rounded consonants with no comparable English sounds, because the inner parts of the lips are rounded and the sound is also palatalized. These sounds include tw [tɕ], dw [dɕ], and hw [ɕ]. The syllabic nasals *m n* (representing both [n] and [ŋ]) always have the same position of articulation as the following consonant, thus *mpaboa* 'shoes' but *nsuo* 'water'.

The most obvious difference between Fante and the other dialects is that in Fante, *t* and *d* are pronounced [ts] and [dz] before front vowels. Thus Fante has *dzi*, meaning ‘eat’, whereas other dialects have *di*, and *itsir* ‘head’, whereas other dialects have *etire* (or *eti*). Also before front vowels, *n* in Fante is pronounced as *ny*; for example, *nye* ‘and’ is *ne* in other dialects. The sound [l] occurs mainly in loanwords from English, although it exists in both Asante and Fante dialects as an alternative pronunciation for [r] or [d] in some words.

Vowels and Vowel Harmony Akan has nine oral vowel phonemes, /i ɪ e ɛ u ʊ o ɔ a/, and five nasal vowels, /ĩ ĩ̃ ã ã̃/. The vowels [ɪ] and [ʊ] are spelled *e* and *o*, respectively. Asante and Akuapem have a tenth vowel, [ɑ]. These vowels pattern according to the rules of cross-height or advanced tongue root vowel harmony. This means that any of the vowels except [ɑ] can be the vowel of a stem syllable, but for prefixes and some suffixes the vowels fall into two sets. These are /a ɪ e ʊ ɔ ĩ ã̃/ and /ɑ ɪ e u o ĩ̃ ã̃/. A prefix to a word must have a vowel from the same set as the stem vowel. Thus, for example, the pronoun prefix meaning ‘he, she’ is pronounced [o] in *odi* ‘she eats’, but [ɔ] in *ɔhwɛ* ‘he looks at it’, because the verb stem vowels /i/ and /ɛ/ belong to different sets.

The Fante dialects also have rounding harmony, whereby the prefix vowels are rounded if the stem vowel is. Thus, in Fante, the expression meaning ‘I am going’ is pronounced [mv-rv-kɔ], because the stem *kɔ* has a rounded vowel, but in other dialects it is pronounced [mi-ri-kɔ].

Tone Every syllable carries contrastive tone. There are two contrastive tone levels, high and low. In a sentence or phrase the pitch of high tones is lowered after a low tone, so that in a sentence such as *Pàpá Kòfí rèfré nè bá* ‘Papa Kofi is calling his child’, each high tone syllable is pronounced at a lower pitch than the earlier high tone syllables. Tone is not reflected in any of the Akan orthographies.

Word Formation

Nouns Most nouns consist of a stem with a singular or a plural prefix. The common singular prefixes are created using the vowels *o*, *e*, and *a* (varying according to the vowel harmony rules), and the common plural prefixes use the vowel *a* (only if there is a different vowel prefix in the singular) or a syllabic nasal. Thus we have *ɔ-hene* ‘king’, plural *a-hene*, and *ɔ-kwasea* ‘fool’, plural *n-kwasea*. Some nouns have no singular prefix, only a plural: thus *gyata* ‘lion’, plural *a-gyata*, and *kuku* ‘pot’, *n-kuku* ‘pots’. Some

adjectives also have singular and plural forms, but there is no noun class agreement of the Bantu type.

Nouns referring to persons often have a suffix *-ni* in the singular, which is replaced by *-fo* in the plural. Thus, *o-buro-ni* ‘European person’, in the plural is *a-buro-fo*. Kinship terms are usually formed with a suffix *-nom* with no change in the prefix, e.g., *ɛna* ‘mother’, *ɛna-nom* ‘mothers’.

Verbs With slight variations among the dialects, the Akan verb is inflected principally for aspect: completive with a suffix with a form that depends on the final stem vowel, perfect with the prefix *á-*, progressive with the prefix *re-*, and habitual and stative forms that have no prefix or suffix and differ only in the tone of the verb. There is also a future marker *bé-*. The consecutive form has a prefix *a-* and is used only in serial verb constructions. The negative is expressed by a prefix consisting of a syllabic nasal before the verb stem, and the imperative also by a syllabic nasal prefix but with high tone.

Syntax

Word Order Akan has subject-verb-object word order. In a noun phrase, adjectives and determiners follow the noun but possessives precede it, as shown in the following examples:

Abofra no re- n- noa bi
child the PROG-NEG-cook some
‘the child will not cook any’

Kwasi kye-ɛ abofra no paanoo
Kwasi give-COMPL child the bread
‘Kwasi gave the child bread’

Amma sika
‘Amma’s money’

Postpositions Locations are represented by a special class of nouns called postpositions at the end of the locative phrase. An example is *so* ‘top, on’, as in the following sentence:

Sekan bi da ɔpon no so
knife some lie table the on
‘a knife is lying on the table’

There is only one preposition, *wɔ* ‘at’.

Serial Constructions Serial verb constructions, in which two or more verbs and their objects occur in sequence with a single subject and no conjunctions to form a complex clause, are a characteristic feature of Akan syntax. For example:

Kwasi de paanoo kye-ɛ abofra no
Kwasi took bread give-COMPL child the
‘Kwasi gave bread to the child’

ɔ-bɛ-tɔ nwoma no a-kan
 she-FUT-buy book the CONSEC-read
 'she will buy the book and read it'

Bibliography

- Dolphyne F (1988). 'The Central Comoé (Tano) languages.' In Dakubu M E Kropp (ed.) *The languages of Ghana*. London: Kegan Paul International. 50–76.
- Dolphyne F A (1988). *The Akan (Twi-Fante) language, its sound systems and tonal structure*. Accra: Ghana Universities Press.
- Dolphyne F A (1996). *A comprehensive course in Twi (Asante) for the non-Twi learner*. Accra: Ghana Universities Press.

- Osam E K (2004). *The Trondheim lectures—an introduction to the structure of Akan: its verbal and multi-verbal systems*. Legon: University of Ghana Department of Linguistics.
- Owusu-Sarpong C (2000). *La mort akan, étude ethno-sémiotique des textes funéraires akan*. Paris: L'Harmattan.
- Yankah K (1989). *The proverb in the context of Akan rhetoric: a theory of proverb praxis*. New York: Peter Lang.
- Yankah K (1995). *Speaking for the chief, okyeame and the politics of Akan royal oratory*. Bloomington and Indianapolis: Indiana University Press.

Akkadian

G Deutscher, Leiden University, Leiden,
The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Akkadian is an extinct Semitic language spoken in ancient Mesopotamia, the 'land between the rivers' (Tigris and Euphrates), in an area that roughly corresponds to today's Iraq. In the later second millennium B.C., Akkadian was also a lingua franca throughout the Near East. Akkadian was written on clay tablets in the cuneiform script in a system that combined syllabic and logographic signs. It is one of the earliest and longest attested languages, with a history that starts around 2500 B.C. and spans more than two thousand years. The ancient name of the language, Akkadûm, derives from the city of Akkade, founded by King Sargon as his capital around 2300 B.C.

From the second millennium B.C., two distinct dialects of Akkadian emerged: Babylonian and Assyrian. Babylonian was spoken in the southern part of Mesopotamia, and Assyrian was spoken in the northern part. During the first millennium B.C., Aramaic gradually ousted Akkadian as the language of the region, and Akkadian ceased to be spoken sometime around 500 B.C. Some texts in Akkadian continued to be written even until the first century A.D., but the language then fell into oblivion, and was rediscovered only in the nineteenth century, when the cuneiform writing system was deciphered. Today, hundreds of thousands of Akkadian texts have been discovered, encompassing many different genres, including poetry (such as the epic of Gilgamesh),

religious compositions, royal and monumental inscriptions, histories, monolingual and multilingual dictionaries (word-lists), grammatical texts, astronomical and mathematical texts, legal documents (such as the Code of Hammurabi), private and diplomatic correspondence, and an endless quantity of economic and administrative documents.

The history of the Akkadian language is conventionally divided into four main chronological periods: Old Akkadian (2500–2000 B.C.), Old Babylonian/Old Assyrian (2000–1500 B.C.), Middle Babylonian/Middle Assyrian (1500–1000 B.C.), and Neo-Babylonian/Neo-Assyrian (1000–500 B.C.). The conventional name 'Old Akkadian' for the earliest attested period is based on the (probably mistaken) assumption that no dialectal variation between the Babylonian and Assyrian idioms existed before the second millennium. The Old Babylonian dialect was considered the classical stage of the language by later generations of Babylonians and Assyrians, and it was the language towards which the later literary idiom (sometimes known as 'Standard Babylonian') aspired.

Grammatical Sketch

During the third millennium B.C., speakers of Akkadian were in prolonged and intimate contact with speakers of the unrelated and typologically dissimilar Sumerian (ergative, agglutinating, verb-final). In consequence, the structure of Akkadian shows an interesting mixture between inherited Semitic features (nominative-accusative alignment, synthetic non-concatenating morphology, noun-modifier order in the NP) with features acquired through convergence.

Such ‘Sprachbund’ effects are evident especially in the phonology and the syntax, as well as in massive lexical borrowing.

The phonemic system of Akkadian underwent a considerable reduction from the putative Proto-Semitic inventory, with the loss of most of the laryngeal and pharyngeal consonants, probably because of contact with Sumerian. Morphology is the area which shows the least evidence of convergence (although even here, some features, such as the ‘ventive’ suffix *-am* may be due to Sumerian influence). Nouns have two genders (masculine, feminine), three cases (nominative, accusative, genitive), and show a distinction between singular, plural, and a partly productive dual.

As in the other Semitic languages, verbal morphology is highly synthetic, and based on a system of mostly three-consonantal roots and internal vowel patterns, combined with prefixing, suffixing, infixing, and gemination. The root **p-r-s** ‘cut’, for instance, appears in forms such as *i-prus* (3SG-cut.PAST), *purs-ā* (cut.IMPERATIVE-2PL), *a-parras* (1SG-cut-NON PAST), *pars-at* (cut.STATIVE-3FSG), *i-pparis* (3SG-cut.PAST.PASSIVE), *nu-šapras* (1.PL-cut.NON PAST.CAUSATIVE).

Where Akkadian morphology diverges significantly from the other (and later attested) Semitic languages, especially in its so called ‘stative conjugation,’ Akkadian seems to present an earlier situation. The ‘stative’ has its origin in conjugated forms of the predicative adjective, but it gradually acquired verbal features. In Akkadian, the stative had not yet become a fully verbal form, but in the other Semitic languages, it was fully integrated in the verbal paradigm (as the ‘perfect’), and this led to a restructuring in the tense-aspect system. The morphology of Akkadian remained fairly stable until the first millennium B.C., when the weakening and loss of final syllables led to the disintegration of the case system on nouns, and to the loss of some distinctions on verbs, and so to the appearance of more periphrastic constructions.

Akkadian is nominative-accusative in both morphology and syntax, and generally has dependent marking, although the verb has obligatory subject agreement as well as direct and indirect object pronominal suffixes. Akkadian word order is interesting, because it can be considered highly ‘inconsistent.’ Akkadian must have inherited a VSO word-order

from Proto-Semitic, and this order is still reflected in archaizing personal-names, especially from the earliest period, such as *Iddin-Sin* (gave:3MSG-Sin – ‘(the god) Sin gave’).

However, undoubtedly because of contact with Sumerian, Akkadian acquired a strict verb-final word order, which is attested from the earliest documents. Both SOV and OSV orders are common, but the only constituents that can follow the verb are the bound object pronoun suffixes (and in later periods finite complement clauses). Nevertheless, inside the noun phrase, Akkadian has retained the characteristic Semitic ‘VO’ characteristics: prepositions, Noun-Genitive, Noun-Relative, Noun-Demonstrative, Noun-Adjective orders. These apparently inconsistent word-order patterns showed no signs of instability, and were maintained intact for two thousand years.

Sources

An extensive state-of-the art overview and bibliography is Huehnergard and Woods (2004). The standard reference grammar is von Soden (1995); Huehnergard (1997) is a teaching grammar. The two research dictionaries are the encyclopaedic Gelb *et al.* (1956-), and von Soden, (1965–1981). Black *et al.* (1999) is a definitions-only dictionary with the most up-to-date overview of the Akkadian lexicon.

Bibliography

- Black J, George A & Postgate N (1999). *A concise dictionary of Akkadian*. Wiesbaden: Harrassowitz.
- Gelb I J *et al.* (eds.) (1956-). *The Assyrian dictionary of the Oriental Institute of the University of Chicago* (21 vols). Chicago: Oriental Institute of the University of Chicago.
- Huehnergard J (1997). *A grammar of Akkadian*. Atlanta: Scholars Press.
- Huehnergard J & Woods C (2004). ‘Akkadian and Eblaite.’ In Woodard R D (ed.) *The Cambridge encyclopedia of the world’s ancient languages*. Cambridge: Cambridge University Press. 218–287.
- von Soden W (1965–1981). *Akkadisches Handwörterbuch* (3 vols). Wiesbaden: Harrassowitz.
- von Soden W (1995). *Grundriss der akkadischen Grammatik* (3rd edn.). Roma: Pontificium Institutum Biblicum.

Albanian

B Demiraj, University of Munich, Munich, Germany

A Esposito, Oxford English Dictionary, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Linguistic Type

Albanian constitutes a single branch of the Indo-European family of languages. It is often held to be related to Illyrian, a poorly attested language spoken in the western Balkans in classical times, but this has not yet been proved conclusively. Although as a people the Albanians have been known since the 2nd century A.D., the earliest surviving records of the Albanian language date only from the 15th century. In its grammar Albanian displays several characteristic features of Indo-European languages, such as declension of nouns by means of case endings and conjugation of verbs by means of personal endings; in its lexicon it preserves a considerable number of words of inherited Indo-European stock.

Albanian may further be characterized as a member of the Balkan Sprachbund. During the many centuries of their evolution the languages of the Balkans (several languages not directly related and belonging to different branches of Indo-European) have come to share certain linguistic features with each other that they do not share with other non-Balkan languages to which they are ostensibly more closely related. Albanian displays several of these features, for example: postposition of the definite article, analytic formation of the future tense (in Albanian with the semiauxiliary verb *dua* 'to want' in the fossil form *do*), substitution of the infinitive by subjunctive clauses, pronominal doubling of objects.

In addition to features shared respectively with other Indo-European languages and with other Balkan languages, Albanian also displays several innovative features, in phonology, in morphosyntax, and in lexis, which mark it out from other European languages.

The phonemic inventory of standard Albanian comprises 7 vowels and 29 consonants, and is remarkable for the way that phonetically similar consonants (including plosives, affricates, fricatives, and liquids) have formed phonemic pairs. The phonological system also reveals the operation of umlaut in former times (with which compare the Germanic languages). As regards morphosyntactic structure, may be mentioned the development, alongside the postpositive definite article, of a proclitic article with indefinite function, which, in turn, has given rise to further innovations: the creation of a special class of adjectives and the reformation of ordinal numerals and of the genitive case. Another important

innovation is the development of the admirative mood in the verbal system, used to express surprise, disagreement, etc.

Present-day Albanian may be categorized as a partly synthetic, partly analytic language, which, alongside synthetic features (both inherited and innovatory), has also developed several analytic features, such as the formation of the perfect and future tenses with auxiliary verbs and the frequent use of prepositions with inflected forms of nouns and pronouns.

The vocabulary of Albanian is notable for the high level of borrowing it shows from different neighboring and influential languages over the course of many centuries, for example: ancient Greek and Latin, the Slavic languages of the Balkans, Turkish, medieval and modern Greek, and (in our own times) French, Italian, and English.

Geographic Spread

Today Albanian is spoken by a population of about 6 500 000 native speakers in a compact ethno-linguistic area in the western Balkans, which comprises:

1. Albania;
2. almost the whole of Kosovo;
3. a broad band of northwestern Macedonia (the former Yugoslav republic) from Kumanovo to Struga;
4. the districts of Medveda, Preševo, and Bujanovac in southern Serbia;
5. the southern and southwestern part of Montenegro;
6. the region of Chameria in northwestern Greece.

Albanian is the official language of the Republic of Albania, and one of the official languages of Kosovo (U.N. administration) and the Republic of Macedonia; it is a national minority language in the Republic of Montenegro.

Outside this compact ethno-linguistic area Albanian is also spoken today in a considerable number of linguistic pockets in the Balkans and beyond. These have arisen as a result of continuing economic and political migrations over the last 700 years. The descendants of the earliest attested diaspora of Albanian-speakers live in scattered communities in southern Greece (the Peloponnese, Attica, and the Aegean islands); the original migration dates from the 14th and 15th centuries, and its cause appears to have been chiefly economic (see Jochalas, 1971). Further scattered communities of Albanian-speakers are to be found in southern Italy and Sicily, where their ancestors settled during the 15th and 16th centuries for political and religious reasons after the

occupation of the western and southern Balkans by the Ottoman Turks. The exact number of Albanian-speakers in these linguistic pockets is difficult to determine, as many of them, especially the younger generation, have abandoned their ancestral language, and speak Greek or Italian, respectively. Those who still retain Albanian (all of whom are bilingual) speak an archaic variety heavily influenced by the superstrate language.

Other linguistic pockets, which, however, are now in danger of being completely assimilated, exist in Serbia (the Sanjak), Croatia (Zadar), central Macedonia, south-eastern Bulgaria (Mandrica), Turkey, and the Ukraine.

During the 20th century emigration of Albanian speakers has continued, especially at the beginning and end of the century from Albania to the United States, Canada, Italy, Greece, and the United Kingdom, and from Yugoslavia (and its successor states) and northern Greece to Turkey, Germany, Switzerland, and Sweden.

Dialects

Within the compact ethno-linguistic area in the western and central Balkans, Albanian is spoken in two main dialects, Gheg and Tosk, each of which may be further divided into several subvarieties. The River Shkumbin in central Albania historically forms the boundary between these two dialects, with the population to the north speaking varieties of Gheg and the population to the south varieties of Tosk (see Gjinari, 1989).

Gheg and Tosk are distinguished from one another chiefly by several important phonological developments. For example, in Tosk /a/ before a nasal has become a central vowel (schwa), and intervocalic /n/ has become /r/. These two sound changes have affected only the old pre-Slav stratum of the Albanian lexicon, that is, native words and loanwords from ancient Greek and Latin. The only important dialectal difference in grammatical structure is the loss of the infinitive in Tosk, in which constructions with the subjunctive predominate just as in all other Balkan languages (with the exception of Serbian and Croatian). However, these innovations, as those that are also evident in different varieties of Gheg, are not such as to impede communication between speakers of the two dialects. Furthermore, the major part of the Albanian lexicon is common to the two dialects.

Of the two main varieties of Albanian spoken outside the ethno-linguistic area, Arvanitika (spoken by the descendants of the ancient migration to Greece) and Arbëresh (spoken by the descendants of the ancient migration to Italy), both preserve archaic features characteristic of varieties of southern Tosk. (The majority of emigrants in these historical

migrations were from southern Albania.) The archaic dialectal features and the separate development of these varieties under the powerful influence of superstrate languages (Greek and Italian) make communication between speakers of the diaspora and those of the ethno-linguistic homeland almost impossible. This differentiation, conditioned by time and space, has caused several specialists to treat these varieties as separate languages (see Sasse, 1991).

Overlying the dialectal diversity of Albanian are different religious (Catholic, Orthodox, Muslim), cultural, and political allegiances that over time have also greatly influenced linguistic developments.

Codification

Up until the early 20th century Albanian was written in a variety of scripts (Roman, Greek, Arabic, Cyrillic), depending on local influences. In 1908 the Congress of Monastir decided on the adoption of the Roman alphabet. The use of Albanian as an official language first became possible after the proclamation of independence of Albania in 1912. However, the emergence of an agreed standard language took time; competing local standards continued to be used until well into the second half of the 20th century. Modern standard Albanian (largely Tosk-based), which is today the accepted standard throughout the whole ethno-linguistic area, did not gain its final sanctioning until 1972 at the Orthographic Congress of Tirana, organized by the Albanian Academy of Sciences, in which linguists and writers from Yugoslavia and the Albanian diaspora also participated.

Present and Future Trends

The decade of the 1990s saw great upheavals in the western Balkans (the fall of communism in Albania, the dismemberment of Yugoslavia, and the war in Kosovo) that radically affected the lives of Albanian speakers. One consequence has been a dramatic increase in the influence of foreign languages on Albanian. A flood of loanwords, especially from English and Italian, is pouring into both the colloquial and the standard language. There exists an unofficial movement opposed to the use of 'unnecessary' foreign words, but attempts to engage the interest of the state in support of its efforts have so far proved unsuccessful.

Bibliography

Camaj M (1984). *Albanian grammar*. Wiesbaden: Otto Harrassowitz.

- Demiraj S (1988). *Gjuha shqipe dhe historia e saj*. Tirana: Shtëpia botuese e librit universitar.
- Demiraj S (1998). 'Albanian.' In Giacalone Ramat A & Ramat P (eds.) *The Indo-European languages*. London: Routledge.
- Gjinari J (1971). 'De la continuation de l'Illyrien en Albanais.' In Korkuti M, Anamali S & Gjinari J (eds.) *Les Illyriens et la genèse des Albanais*. Tirana: Université de Tirana. 173–181.
- Gjinari J (1989). *Dialektet e gjuhës shqipe*. Tirana: Akademia e shkencave të republikës popullore socialiste të Shqipërisë.
- Hamp E P (1972). 'Albanian.' In Sebeok T A (ed.) *Current trends in linguistics*, 9. The Hague: Mouton. 1626–1692.
- Jochalas T (1971). 'Über die Einwanderung der Albaner in Griechenland; eine zusammenfassende Betrachtung.' In *Dissertationes Albanicae: in honorem Josephi Valentini et Ernesti Koliqi septuagenariorum*. Munich: Trofenik.
- Hetzer A (1985). *Geschichte des Buchhandels in Albanien*. Wiesbaden: Otto Harrassowitz.
- Hetzer A (1996). 'Die vereinheitliche albanische Schriftsprache.' In Schaller H (ed.) *Sprache und Politik: die Balkansprachen in Vergangenheit und Gegenwart*. Munich: Südosteuropa-Gesellschaft.
- Newmark L (1998). *Albanian–English dictionary*. Oxford: Oxford University Press.
- Raka F (1997). *Historia e shqipës letrare*. Prishtina: Rilindja.
- Sasse H J (1991). *Arvanitika*. Wiesbaden: Otto Harrassowitz.

Algonquian and Ritwan Languages

D H Pentland, University of Manitoba, Winnipeg, MB, Canada

© 2006 Elsevier Ltd. All rights reserved.

More than 30 languages of the Algonquian family were formerly spoken along the east coast of North America from about 34°N (Cape Fear, North Carolina) to about 56°N (Davis Inlet, Labrador), around the upper Great Lakes, and west to the foothills of the Rocky Mountains. They were the first North American languages encountered by French and English explorers; by the end of the 17th century several languages had already been described in detail. Three centuries later, however, two-thirds of the languages are no longer spoken, with only English loanwords such as *moccasin*, *skunk*, and *squaw* to reflect their former existence. The 'Ritwan' languages (Wiyot and Yurok) of California are distantly related. Pilling (1891) provides a nearly exhaustive inventory of the earlier sources; later publications are listed by Pentland and Wolfart (1982), but the only comprehensive bibliography of the most recent literature is in Nichols (1981–).

Classification

The only widely accepted genetic subgroup within the Algonquian family is Eastern Algonquian, consisting of the languages which descended from Proto-Eastern Algonquian (Goddard, 1978b). It includes the languages of the Maritime provinces, southern Quebec, and the northern New England states – Micmac (several dialects), Malecite-Passamaquoddy, Etchemin, Eastern and Western Abnaki (two languages, each

with several dialects), and Pocumtuck or 'Loup B' – and those formerly spoken in the Hudson and Delaware River basins of New York, Pennsylvania, and New Jersey – two dialects of Mahican, and the two 'Delaware' languages, Munsee (including the divergent Wappinger dialect) and Unami (three dialects). The languages of southern New England and Long Island – Nipmuck ('Loup A'), Massachusetts (Wampanoag), Narragansett, Pequot-Mohegan-Montauk, and Quiripi-Unquachog – and those of the southeastern states – Nanticoke, Conoy (Piscataway), Powhatan (Virginia Algonquian), and Roanoke-Pamlico (Carolina Algonquian) – may also be part of the Eastern subgroup, but since all are extinct, the crucial phonological details depend on interpretations of early written records.

The so-called 'Central' languages were located between Hudson Bay and the Ohio River valley; each shares many features with its neighbors, but there are no ancient subdivisions.

Cree-Montagnais-Naskapi is a dialect chain extending across central Canada from Labrador to Alberta, conventionally subdivided according to the reflex of Proto-Algonquian **l*: Plains Cree (Nêhiyawêwin), the dialect with *y* < **l*, in Alberta and Saskatchewan; three varieties of Woods Cree (with *ð*) in northern Saskatchewan and Manitoba, one of which probably continues the extinct Missinipi dialect (with *r*; cf. Pentland, 2003); three or more varieties of Swampy Cree (with *n*) in Manitoba and northern Ontario; Moose Cree (with *l*) on the southwest coast of James Bay; and Atikamekw (or Tête de Boule, with *r*), in southwestern Quebec, cut off from the others by a dialect of Ojibwa. In the eastern dialects Proto-Algonquian **k* has

palatalized to *č* before front vowels: Eastern Montagnais (Innu-aimun) and Eastern Naskapi (with *n* < **l*), in Labrador and southeastern Quebec; Southern Montagnais (with *l*), at Lac St-Jean, Quebec; and the extinct dialect of Tadoussac, Quebec (with *r*). The several varieties of East Cree and Western Naskapi in northern Quebec (all with *č* < **k* and *y* < **l*) are considered transitional between the eastern and western dialects (MacKenzie, 1980), or as varieties of a Western Montagnais dialect (Pentland, 1978); some East Cree speakers understand Moose Cree, but speakers of the nonpalatalized dialects generally find East Cree and (other) Montagnais dialects completely unintelligible.

Ojibwa (also spelled Ojibway or Ojibwe) is another dialect chain, extending from Quebec to Saskatchewan. The Algonquin dialect of southwestern Quebec is separated by a large number of isoglosses from its immediate neighbors (Rhodes and Todd, 1981), but shares a number of features with Northern (or Severn) Ojibwa, in northwestern Ontario. A quite different dialect, also usually called Algonquin, is spoken at Maniwaki, Quebec; it apparently is the result of a large migration of Eastern Ojibwa speakers from Lake Nipissing into an originally Algonquin-speaking community at Oka. The Eastern Ojibwa dialect of southern Ontario and the Ottawa (or Odawa) dialect of Michigan and southwestern Ontario have both reduced or lost all unstressed vowels. According to Rhodes and Todd (1981), the other dialects are Central Ojibwa, in northeastern Ontario; Northwestern Ojibwa, between Lake Superior and Lake Winnipeg; Southwestern Ojibwa (Chippewa), in northern Michigan, Wisconsin, and Minnesota; and Saulteaux (Plains Ojibwa) in southern Manitoba and eastern Saskatchewan.

Potawatomi, originally spoken in southern Michigan, was once a part of the Ojibwa dialect chain; it separated before Ojibwa merged **ye* with *i*; prior to the first contact with Europeans, but shares with some southern Ojibwa dialects the complete loss of unstressed vowels. Menomini (or Menominee), in Wisconsin, has many Ojibwa loanwords and shares some sound changes (including **ye* > *i*), but is in other respects quite different from other Algonquian languages.

Four dialects of a single language were formerly spoken in southern Michigan: Fox (or Mesquakie), Sauk, Kickapoo, and the extinct Mascouten dialect. The three surviving varieties are probably still mutually intelligible, but Kickapoo has some significant differences.

The states of Illinois and Indiana were the home of the Miami-Illinois language, which contained

a number of dialects, including Kaskaskia, Peoria, Tamaroa, Wea, Piankashaw, and Miami; by the 1870s there were only two groups, known as Peoria and Miami, but they may not correspond to older dialect divisions. In the early 18th century the Michigamea spoke a dialect of Illinois (cf. Masthay, 2002: 26), but earlier may have spoken an unrelated language (Goddard, 1978a: 587).

The Shawnee originally lived in southern Ohio, but during the historic period they often split into widely scattered bands, eventually merging into three politically independent groups, the Eastern Shawnee, Cherokee Shawnee, and the Absentee Shawnee, all now resident in Oklahoma. Neither early nor recent dialect differences have yet been examined in detail.

In addition to Plains Cree and Plains Ojibwa (Saulteaux), there were at least six other Algonquian languages spoken on the Great Plains (Goddard, 2001). Blackfoot is spoken in Alberta by the Blackfoot (Siksika), Blood, and Northern Peigan, and in Montana by the Southern Peigan (or Blackfeet) with only slight differences. Arapaho (including the extinct Besawunena dialect, in Wyoming and Oklahoma) is closely related to Atsina or Gros Ventre (in Montana). Some Arapaho formerly spoke Ha'anahawunena, an unrecorded language said to have been very different from Arapaho; the Southern Arapaho originally spoke Nawathinehena, a distinct Algonquian language of which only a few words were recorded in 1899.

The two modern Cheyenne communities in Montana and Oklahoma speak almost identical dialects; the Sutaio, who joined the Cheyenne in the 19th century, spoke a different dialect or language, but little reliable information about it was ever recorded.

In 1913 Edward Sapir showed that Wiyot and Yurok, two languages of northwestern California which had just been assigned to a new linguistic family called Ritwan, are related to the Algonquian languages.

Sapir extended the name Algonkin (i.e., Algonquian) to the larger group. This unfortunate relabeling was misunderstood by Truman Michelson, who argued (correctly) that Wiyot and Yurok are not 'Algonquian' in the same sense as Fox or Cree; he was wrong, however, to deny the more distant relationship, which later work has amply confirmed.

The family consisting of the Algonquian languages plus Wiyot and Yurok is now called Algic; the name Ritwan is reserved for Wiyot and Yurok, should it turn out that they form a single branch within the Algic family: the question is still undecided. The last speaker of Wiyot died in 1962; fieldwork continues with the last few speakers of Yurok.

The extinct Beothuk language of Newfoundland may have been related to the Algonquian family, but the early 19th-century vocabularies are poorly transcribed and very inconsistent (Hewson, 1978); some words and inflections appear to be cognate, but others bear no resemblance to their Algonquian counterparts, even allowing for the usual kinds of transcription errors. It is unlikely that the relationship (if there is one) will ever be demonstrated satisfactorily.

Edward Sapir placed Algonquian in a stock with Kutenai and the Salishan, Chimakuan, and Wakashan families, but the similarities he noted are probably ancient loans or areal features. The few resemblances between single morphemes in Proto-Algonquian and the languages of the Gulf coast are probably coincidental.

Demography

No accurate census of Algonquian speakers exists. According to the 2001 Canadian census there were 72 680 Cree people, but 102 185 speakers of the Cree language; Grimes (1992) estimated 42 725 speakers, but even this number may be too high. An additional 14 000 people speak the ‘palatalized’ dialects, East Cree, Naskapi, and Montagnais.

There are at least 20 890 speakers of Ojibwa in Canada (2001 census) and perhaps 30 000 in all; earlier estimates ranged above 50 000 speakers. About 40–50 fluent speakers of Potawatomi remain, although 200–500 were estimated 30 years ago. A few dozen elderly people still speak Menomini. Perhaps 200 people still speak Fox or Sauk, but Kickapoo has well over 1000 speakers. Shawnee is said to have 200–250 speakers; the Miami-Illinois language became extinct about 50 years ago.

Of the Eastern Algonquian languages, only Micmac and Malecite-Passamaquoddy are still viable. There may be as many as 8000 speakers of Micmac in the Maritime provinces and southern Quebec, and more than 1000 speakers of Malecite-Passamaquoddy in New Brunswick and Maine. The last speaker of Penobscot (Eastern Abenaki) died in 1993; a few elderly people may still speak Western Abnaki. Perhaps a dozen people in southern Ontario speak Munsee Delaware, but Unami, in Oklahoma, is virtually extinct.

The 2001 Canadian census reported 2740 Blackfoot in Canada, but 4495 speakers of the language; there may be 5000 speakers in all, including a few children. Arapaho is estimated to have several hundred fluent speakers (Goddard, 2001), but there are only two speakers of Atsina (Gros Ventre) left. Cheyenne is spoken by about 2500 people.

Since the number of speakers of many Algonquian languages has declined rapidly in recent years, many communities have sought to revitalize their traditional language by introducing language programs in the local schools. A few programs have been very successful, but many others have failed to increase the use of the language outside the classroom.

Recent attempts to revive extinct languages such as Miami-Illinois and Pequot-Mohegan cannot yet be evaluated.

Typological Characteristics

Algonquian languages are polysynthetic, hierarchical, nonconfigurational head-marking languages with discontinuous constituents and relatively free word order.

Phonology

The parent language, Proto-Algonquian (PA), was reconstructed by Leonard Bloomfield (1925, 1946), in part to demonstrate that the comparative method can be applied successfully to ‘unwritten’ languages as well as those with ancient records. PA probably had 13 consonants (**p*, *t*, *k*, *kʷ*, *s*, *š*, *h*, *θ*, *l*, *m*, *n*, *w*, *y*) and four short and four long vowels (**a*, *e*, *i*, *o*; **aː*, *eː*, *iː*, *oː*). Bloomfield also reconstructed **č*, but it occurs only before **i*(·) and **y* (where it does not contrast with **t*); however, **č* may also have replaced **t* in words with diminutive consonant symbolism. He did not reconstruct **kʷ*, but it probably contrasted with the sequence **kw*. Consonant clusters could not occur word initially, and every word ended in a vowel (usually, but not always, a short vowel).

In PA, stress was predictable, with all long vowels and every second short vowel receiving a stress; this stress system is preserved with little change in Ojibwa, and underlies the vowel length alternations in Menomini, but some languages (e.g., Plains Cree, Montagnais) have replaced it with systems which count syllables from the end of the word, and Miami-Illinois reflects both types. Arapaho-Atsina and Cheyenne have developed pitch accent systems (largely from the old length contrast), while others (Eastern Montagnais, Kickapoo, and Malecite-Passamaquoddy) have acquired pitch contrasts from the loss or contraction of certain syllables.

Almost all the daughter languages have merged PA **θ* and **l*, and some have a further merger with **n* (as in Massachusetts and in modern Ojibwa, Menomini, and Fox) or with **y* (as in Pequot-Mohegan); although the PA phonetic values of the consonants Bloomfield labeled **θ* and **l* are debated, the reflexes in Table 1 clearly show that they

Table 1 Intervocalic reflexes of five Proto-Algonquian consonants in selected languages

| | *t | θ | n | l | y |
|----------------|----|---|---|---|---|
| Plains Cree | t | t | n | y | y |
| Swampy Cree | t | t | n | n | y |
| Ojibwa, Fox | t | n | n | n | y |
| Shawnee | t | l | n | l | y |
| Pequot-Mohegan | t | y | n | y | y |
| Arapaho | t | θ | n | n | n |

were distinct phonemes in PA. In morpheme-final position, *t and *θ still contrast in Cree, and *θ and *l still contrast in Shawnee.

Inflectional Morphology

Nouns are classified as animate (NA) or inanimate (NI), the animate category including not only all living things but also some plants and their products, a few body parts, and miscellaneous other items such as snow, kettles, and snowshoes; all other nominals, including most body parts and the personal pronouns, are grammatically inanimate.

Possession is indicated by a pronominal prefix; plurality of the possessor is marked by a suffix. Most kinship terms and body parts, and a very few other noun stems, are 'dependent' (inalienably possessed); a special 'unspecified possessor' prefix is used with body part nouns when there is no actual possessor (e.g., **me-sit-i* 'someone's foot'), but to express 'a daughter' Algonquian languages must resort to a verbal derivative, literally '(one that) someone has as a daughter.'

Nominals are obligatorily specified as singular (PA *-a NA, *-i NI) or plural (PA *-aki NA, *-ali NI), but with the loss of final vowels singulars have no overt marking in most of the daughter languages. The third person distinguishes between proximate (central, in focus) and obviative, but only animate nouns have separate obviative inflections (PA *-ali obv. sg., *-ahi obv. pl.); otherwise, obviation is evident only in verb agreement. Some languages have a second set of endings to indicate inaccessibility or absence (PA *-a NA sg., *-e NI sg., etc.).

The vocative has distinct singular and plural suffixes. A locative (in *-*[e]nki*) may be derived from any possessed or unpossessed noun stem (as well as a few other initial elements), but it is an uninflected 'particle' which does not distinguish number or obviation.

Intransitive verbs have distinct stems for animate and inanimate subjects, transitive verbs for animate and inanimate objects: e.g., Cree *kisiso* 'be hot

(ANIM)', *kisite* 'be hot (INAN)', *kisisw* 'heat (ANIM)', *kisisam* 'heat INAN'. Animate intransitive (AI), inanimate intransitive (II), and transitive inanimate (TI) stems have similar inflections; transitive animate (TA) stems have more complicated paradigms, since they may distinguish almost any combination of subject and (animate) object.

Verb inflections are divided into three formally distinct sets of paradigms ('orders'). The PA forms of the basic endings were reconstructed by Bloomfield (1946); Goddard (1979) provided much additional information.

The independent order, used primarily in main clauses, employs the same personal prefixes as possessed nouns, to indicate the highest-ranking argument of the verb (as determined by the hierarchy 2nd person > 1st > unspecified > anim. 3rd > anim. obv. 3rd > inan. 3rd > inan. obv. 3rd) if this is not otherwise marked; suffixes indicate direction (direct when the agent of a TA verb outranks the patient, inverse when the agent is not the highest-ranking argument), plurality and obviation, negation, and various modal categories (Pentland, 1999).

The conjunct and imperative orders employ only suffixes to indicate the same categories, but some forms in the conjunct order (such as participles) also have 'initial change' or ablaut of the vowel of the first syllable of the verb complex (Costa, 1996).

Derivational Morphology

Most Algonquian words can be described as consisting of an initial, an optional medial, and a final, each of which may itself be derived from shorter elements (Goddard, 1990). Roots (unanalyzable initials) are typically adjectival or adverbial rather than nominal or verbal, e.g. **melw* 'good, well' (as in **melwa:kam-yi* II 'be good water, taste good [of a liquid]', **melwapam* TA 'like to look at someone', **melwenk^wam* AI 'sleep well') and **wel* 'properly arranged' (as in **welenam* TI 'arrange something by hand, place something in readiness', **welešam* TI 'cut something to shape'). The final determines the word class; thus beside the TI stem **welešam* (with final *-*[e]šam* 'cut-INAN') there is a corresponding TA stem **welešw* 'cut someone to shape' (with *-*(e)šw* 'cut-ANIM'), and further derivatives **welesamaw* TA 'cut something to shape for someone' (with benefactive final *-*aw*-), **welešama:swi* AI 'cut something to shape for oneself' (with reflexive final *-*[e]swi* added to the benefactive), and **welešama:sowen* NI '(act of) cutting something to shape for oneself' (with noun-final *-*wen* added to the reflexive). The addition of an additional final almost always changes the word class.

Medials are nominal elements incorporated between the initial and final. Some are classifiers, such as **-axk^w* ‘wooden’, **-apeθk-* ‘stone or metal’, and **-apye^k* ‘stringlike’, in the II stems **kenwa^{xk^w}at-* ‘be long [of something wooden]’, **kenwa^{peθk}at-* ‘be long [of a stone or metal object]’ and **kenwa^{pye^k}kat-* ‘be long [of something stringlike]’. Others correspond to the direct object of the English equivalent, such as **-neθk-* ‘hand, arm’ in **kenwineθke-* AI ‘have a long hand or arm’ or **-eθk^wew-* ‘woman’ in **no^{teθk^w}ewe-* AI ‘pursue women’, but noun incorporation is not very productive and does not interact with agreement.

Syntax

As many as four noun phrases may occur in a single clause, but no more than two arguments can be marked on the verb by inflectional affixes (Thomason, 2004). All verbs obligatorily take a subject, and may take an instrumental argument. TA stems obligatorily take an animate object; both AI and TI stems may also take an object, and TA stems may take a second object. Instrumentals, AI objects, and TA second objects may be of either gender.

Word order is very free: almost all permutations of constituents are grammatical. A noun phrase may be discontinuous, with part before the verb and the remainder after (Reinholtz, 1999); in Fox, compound verbs may also be discontinuous, with other parts of a clause inserted between a preverb and the remainder of the verb complex, as in (1):

- (1) *ne-kehke^{nem}-ekw-a* *ni-na* *e-h* = *pwa^{wi}*
 1ST-know-INV-3RD.ANIM.SG I COMP = not-
ke^{ko}·hi -*ašeno-ni*-*ki*
 something be.absent-OBV-3RD.INAN.SG
 ‘he knows that as for me, nothing is missing’
 (Dahlstrom, 1995: 9)

The topic of the subordinate clause, *ni-na* ‘I’, has been raised to the left-hand edge of the clause; the subject of the II verb *ašeno-* ‘be absent’ has been moved into the verb complex following the complementizer clitic *e-h* (which bears the ‘initial change’) and a negative preverb.

In example (1) the topic has also been copied as the direct object of the matrix verb, which is therefore the TA stem *kehke^{nem}-* ‘know someone’ rather than TI *kehke^{netam}-* ‘know something’; subjects and (some) objects can also be copied, and the verb of the subordinate clause may be incorporated into the matrix verb, as in the Fox example in (2):

- (2) *ke^{ki}-ši* = *meko* *yo^{we}*
 2ND-already = EMPH in.the.past
nepow-e^{nem}-ene-pena

die-think-2ND.OBJ-1ST.PL

‘we had thought you were already dead’

(Goddard, 1988: 71)

The preverb of the incorporated clause *ki-ši-nep-* ‘have already died’ has been moved to the preverb position of the matrix clause (where it is followed by an emphatic clitic and an adverb) but semantically still modifies only the lower verb.

Mixed Languages

Blackfoot may be descended from a precontact creole: it has (for the most part) normal Algonquian morphology and cognates of many individual morphemes, but few complete words are reconstructible.

A number of pidgins arose during the contact period, based on Powhatan (Virginia, early 17th century), Unami (New Jersey, 17th century), Cree (Hudson Bay, 18th century), and Ojibwa (Lake Superior, 19th century). An early Micmac–Basque pidgin in Nova Scotia was the source of a few Basque loanwords in modern Micmac, such as *eleke^{wit}* ‘(one who is) king’ < Basque *errege*.

Métchif or Michif, a French–Cree mixed language, is still spoken in some Metis communities in North Dakota, Manitoba, Saskatchewan, and Alberta (Bakker, 1997), and a remarkably similar French–Montagnais mixed language has developed at Betsiamites, Quebec. In these languages, the noun phrase is mainly French lexical items with French phonology and morphology, while the remainder of the clause is Plains Cree or Southern Montagnais.

Philology and Documentation

With more than four centuries of records on various languages available, philological studies have long played a role in Algonquian linguistics. The earlier English sources have been utilized by many scholars, notably in a study of the historical phonology of Powhatan (Siebert, 1975). The early French records have not been as thoroughly studied, but editions of older grammars (e.g., Daviault, 1994) and dictionaries (e.g., Masthay, 2002) have increased interest in the use of older materials to elucidate various details in the development of the modern languages.

One problem with the early sources is that they tend to provide individual words and partial paradigms rather than connected sentences; most early textual material is based on European originals, and was probably translated by the missionaries themselves. One notable exception is the collection of Massachusetts documents edited by Goddard and Bragdon (1988). Since the beginning of the 20th century many texts written or dictated by native

speakers have been published, but many more remain in manuscript.

Grammars and dictionaries of many Algonquian languages have been published, but much remains to be done: syntax is seldom treated at length, and some of the dictionaries are pitifully small. Leonard Bloomfield showed the way with a grammar (1962) and an 11 000-word dictionary (1975) of Menomini; notable later productions are the Montagnais-French dictionary compiled by Lynn Drapeau (1991), with nearly 24 000 entries, and the 1100-page reference grammar of Ojibwa by J. Randolph Valentine (2001). Mithun (1999: 328–337) provides a brief survey of the sources available for each of the languages.

Bibliography

- Bakker P (1997). *A language of our own: the genesis of Michif, the mixed Cree-French language of the Canadian Métis*. New York/Oxford: Oxford University Press.
- Bloomfield L (1925). 'On the sound-system of Central Algonquian.' *Language* 1, 130–156.
- Bloomfield L (1946). 'Algonquian.' In Hoiyer H *et al.* *Linguistic structures of native America*. New York: Viking Fund. 85–129.
- Bloomfield L (1962). *The Menomini language*. New Haven: Yale University Press.
- Bloomfield L (1975). *Menomini lexicon*. Hockett C F (ed.) Milwaukee: Milwaukee Public Museum.
- Costa D J (1996). 'Reconstructing initial change in Algonquian.' *Anthropological Linguistics* 38, 39–72.
- Dahlstrom A (1995). *Topic, focus and other word order problems in Algonquian*. Winnipeg: Voices of Rupert's Land.
- Daviault D (1994). *L'algonquin au XVII^e siècle: une édition critique, analysée et commentée de la grammaire algonquine du Père Louis Nicolas*. Sainte-Foy: Presses de l'Université du Québec.
- Drapeau L (1991). *Dictionnaire montagnais-français*. Sillery: Presses de l'Université du Québec.
- Goddard I (1978a). 'Central Algonquian languages.' In Trigger (ed.), 583–587.
- Goddard I (1978b). 'Eastern Algonquian languages.' In Trigger (ed.), 70–77.
- Goddard I (1979). *Delaware verbal morphology: a descriptive and comparative study*. New York/London: Garland.
- Goddard I (1988). 'Post-transformational stem derivation in Fox.' *Papers and Studies in Contrastive Linguistics* 22, 59–72.
- Goddard I (1990). 'Primary and secondary stem derivation in Algonquian.' *International Journal of American Linguistics* 56, 449–483.
- Goddard I (2001). 'The Algonquian languages of the Plains.' In DeMallie R J (ed.) *Handbook of North American Indians*, 13: *Plains*. Washington: Smithsonian Institution. 71–79.
- Goddard I & Bragdon K J (eds.) (1988). *Native writings in Massachusetts* (2 vols). Philadelphia: American Philosophical Society.
- Grimes B F (ed.) (1992). *Ethnologue: languages of the world* (12th edn.). Dallas: SIL International.
- Hewson J (1978). *Beothuk vocabularies: a comparative study*. St. John's: Newfoundland Museum.
- MacKenzie M E (1980). *Toward a dialectology of Cree-Montagnais-Naskapi*. Unpublished Ph.D. thesis, University of Toronto.
- Masthay C (ed.) (2002). *Kaskaskia Illinois-to-French dictionary*. St. Louis, MO: Carl Masthay.
- Mithun M (1999). *The languages of native North America*. Cambridge: Cambridge University Press.
- Nichols J D (1981–). 'Bibliography: Algonquian.' *Algonquian and Iroquoian Linguistics* 5–(continuing).
- Pentland D H (1978). 'A historical overview of Cree dialects.' In Cowan W (ed.) *Papers of the 9th Algonquian Conference*. Ottawa: Carleton University. 104–126.
- Pentland D H (1999). 'The morphology of the Algonquian independent order.' In Pentland D H (ed.) *Papers of the 30th Algonquian Conference*. Winnipeg: University of Manitoba. 222–266.
- Pentland D H (2003). 'The Missinipi dialect of Cree.' In Wolfart H C (ed.) *Papers of the 34th Algonquian Conference*. Winnipeg: University of Manitoba. 287–301.
- Pentland D H & Wolfart H C (1982). *Bibliography of Algonquian linguistics*. Winnipeg: University of Manitoba Press.
- Pilling J C (1891). *Bibliography of the Algonquian languages*. Washington: Smithsonian Institution.
- Reinholtz C (1999). 'On the characterization of discontinuous constituents: evidence from Swampy Cree.' *International Journal of American Linguistics* 65, 201–227.
- Rhodes R A & Todd E M (1981). 'Subarctic Algonquian languages.' In Helm J (ed.) *Handbook of North American Indians*, 6: *Subarctic*. Washington: Smithsonian Institution. 52–66.
- Siebert F T Jr (1975). 'Resurrecting Virginia Algonquian from the dead: the reconstituted and historical phonology of Powhatan.' In Crawford J M (ed.) *Studies in southeastern Indian languages*. Athens: University of Georgia Press. 285–453.
- Thomason L (2004). 'Two, three and four noun phrases per clause in Meskwaki.' In Wolfart H C (ed.) *Papers of the 35th Algonquian Conference*. Winnipeg: University of Manitoba. 407–30.
- Trigger B G (ed.) (1978). *Handbook of North American Indians*, 15: *Northeast*. Washington: Smithsonian Institution.
- Valentine J R (2001). *Nishnaabemwin reference grammar*. Toronto: University of Toronto Press.

Altaic Languages

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

A common designation for the typologically related languages of the Turkic, Mongolic, and Tungusic families is 'Altaic languages'; according to some scholars, this designation also includes Korean and Japanese. The common typological features of these languages include an agglutinative and exclusively suffixing word structure, sound harmony, verb-final word order, with dependents preceding their head, and use of numerous nonfinite verb constructions.

Altaic as 'Ural-Altaic'

The term 'Altaic' was first used by M. A. Castrén in the middle of the 19th century for a supposed family comprising Finno-Ugric, Samoyedic, Turkic, Mongolic, and Tungusic. This group of languages was later called 'Ural-Altaic.' The Ural-Altaic hypothesis, which was largely based on general typological criteria such as agglutination and vowel harmony, was widely accepted in the 19th century. Later on, this hypothesis was seriously doubted. The works on 'Altaic' languages by W. Schott, M. A. Castrén, J. Grunzel, H. Winkler, and others contain abundant incorrect data. Castrén, however, rejected the purely typological approach and applied linguistic criteria of lexical and morphological comparison. There are not sufficient materials to establish a Ural-Altaic protolanguage.

Scholars of the following period, e.g., J. Németh and J. Deny, who took a more cautious attitude, published detailed works on phonology, word formation, etc. Syntactic typological arguments for the unity of Ural-Altaic were, however, discussed as late as 1962, by Fokos-Fuchs.

Altaic as 'Micro-Altaic'

Scholars such as G. J. Ramstedt and N. Poppe argued for a 'Micro-Altaic' family (Comrie, 1981: 39) that at least consisted of Turkic, Mongolic, and Tungusic, three well-established genealogical groups. Ramstedt is the founder of Altaic linguistics in a scientific sense, though his works contain many problematic details. His introduction to Altaic linguistics was published posthumously (1952–1957). Poppe's contributions to Altaic linguistics are not less important. His comparative phonology, planned as the first part of a comparative grammar, appeared in 1960. An example of phonological correspondences according to

Ramstedt and Poppe is the supposed development of the initial Altaic stop **p-* into Korean *p-* and *ph-*, into Tungusic *p-* (Nanai), *f-* (Manchu), and *h-* (Evenki), into Mongolic **p-* (Proto-Mongolic), *h-* (Middle Mongolian), *f-* (Monguor), and *Ø-* (Buriat, Oirat, Kalmyk, etc.), and into Turkic *h-* (Proto-Turkic, some modern languages) and *Ø-* (most modern languages). Ramstedt's and Poppe's arguments were largely accepted until they were challenged by G. Clauson (1956, 1962). Opponents such as J. Benzing and G. Doerfer expressed doubts even against this Micro-Altaic unit as a valid genealogical family.

Whereas the Altaicists regarded certain similar features as a common heritage from a protolanguage, others claimed that the similarities were the result of contact processes. Thus certain common features in Mongolic and Chuvash could go back to Proto-Altaic or had been borrowed into Mongolic from a language of the Chuvash type. Clauson had criticized the lack of evidence for a common basic vocabulary in Altaic. In his huge work on Turkic and Mongolic loanwords in Iranian, Doerfer (1963–1975) refuted the Altaic etymologies presented by Ramstedt, Poppe, and others, arguing that similarities that can be attributed to general typological principles or to areal diffusion must be excluded from genealogical comparisons.

A possible Altaic unity must have been dissolved about 3000 B.C. The crucial question in Altaic comparative studies is by which methods common elements due to early contacts can be distinguished from elements inherited from a protolanguage. One problem is the scarcity of early data. Whereas Indo-European is attested already in the second millennium B.C., there are no real Turkic sources prior to the 8th century (East Old Turkic inscriptions in the Orkhon valley, Inner Asia). The first Mongolic materials are found in *The secret history of the Mongols* (believed to be written around 1240 A.D., partly based on older materials). The first substantial materials documenting Tungusic emerge centuries later.

The Turkic–Mongolic–Tungusic Relationship

As for the relationship between Turkic and Mongolic, it has been possible to establish a number of convincing sound laws on the basis of words with similar sound shape and content, and to find certain corresponding derivational and grammatical suffixes. The question is how to judge these similarities. The earliest Turkic and Mongolic sources hardly show any common features except for intercultural words

such as *qayan* 'supreme ruler' and *teγri* 'heaven.' Middle Mongolian displays a number of words with similar Turkic equivalents. The few pairs of corresponding words do not, however, relate to the most significant parts of the vocabulary, i.e., numerals, kinship terms, and basic verbs, nouns, and adjectives. A few common elements are found in morphology. On the other hand, it is obvious that later Mongolic languages have converged with Turkic by giving up some old features, e.g., an inclusive vs. exclusive distinction in pronouns and verbs, grammatical gender in verb forms, agreement between the adjectival attribute and its head, and the option of postposed adjectival attributes.

Many similarities may thus be due to contact processes. There were close ties between Turkic and Mongolic as early as the middle of the first millennium B.C. Borrowings in both directions had taken place since early times. With the rise of the Chingisid Empire in the 13th century, many Turkic varieties came under strong Mongolic influence. The impact lasted longer in areas of intensive contact, such as South Siberia and the Kazakh steppes. The lexical influence is particularly strong in Tuvan, Khakas, Altay Turkic, Kirghiz, Kazakh, etc. Look-alikes that occur only in typical contact zones cannot easily be used as evidence for genealogical relatedness.

Mongolic displays early layers of loanwords from several Turkic languages and has developed many structural traits under Turkic influence. Words common to Turkic and Mongolic, e.g., Bulgar-Mongolic correspondences, are regarded by Altaicists as true cognates and by non-Altaicists as Turkic loans in Mongolic. Some scholars consider the possibility that correspondences between Turkic and Mongolic go back to a common adstrate, some 'language X' that might have delivered loans to both groups. Tungusic words considered by Altaicists as Altaic are rather regarded by non-Altaicists as loans from Mongolic in certain contact areas. Similar derivational and grammatical suffixes are very scarce. Mongolic and Tungusic had been in contact for a long time prior to the first documentation of Tungusic. Except for recent Yakut loans in North Tungusic, there are hardly any plausible lexical correspondences between Turkic and Tungusic. In a non-Altaicist perspective, the overall Turkic-Mongolic-Tungusic relationship thus appears to be due to diffusion rather than to genealogical relatedness. According to this view, words common to all groups may have wandered along the path Turkic → Mongolic → Tungusic.

After decades of discussions, the nature of the relationship between the Altaic languages is still controversial. Many common features are the result of

recent contact, often limited to certain languages within the groups. The question is what reliable correspondences remain to justify the recognition of Altaic as a family in the sense of Indo-European or Semitic. There is no consensus as to whether the relatedness is proven, still unproven, or impossible. Some scholars argue that too few features are common to all three groups, and only to these groups. There are clear lexical and morphological parallels between Turkic and Mongolic, and between Mongolic and Tungusic, but not between Turkic and Tungusic. All three groups exhibit a few similar features, e.g., in the forms of personal pronouns, but similarities of this kind are found in different unrelated languages, in the rest of northern Eurasia and elsewhere. Today, however, compared to the 1960s, the fronts between Altaicists and non-Altaicists are not always as rigid. For example, the pronounced non-Altaicist Doerfer, who had criticized the proposed Altaic sound laws as being construed less strictly or even *ad hoc*, has accepted the above-mentioned development of **p*- into Turkic *h*- and Ø-: e.g., **pat* 'horse,' *bat* (Khalaj, etc.), *at* (most Turkic languages). Doerfer expresses his appreciation of the achievements of the Altaicist Ramstedt in the following way: "We must be grateful to the ingenious founder of Altaistics as a science for discovering so many sound laws which are valid to this date" (Doerfer, 1985: 135).

Korean and Japanese

The most controversial point in recent discussions has been whether Korean and Japanese (with the closely related Ryukyuan language) should be regarded as members of an Altaic family. G. J. Ramstedt (1939, 1949) was the first scholar to attempt to prove a remote relationship between Turkic-Mongolic-Tungusic and Korean. Though his comparisons have been heavily criticized in more recent studies, N. Poppe considered Ramstedt to have identified at least 150 incontestable Korean-Tungusic-Mongolic-Turkic cognates.

Japanese has often been taken to consist of an Austronesian substratum and an Altaic superstratum. E. D. Polivanov (1924) argued that it is of hybrid origin, containing both Austronesian elements and continental elements that are also found in Korean and Micro-Altaic. In an early study, Ramstedt (1924) investigated possible links between Japanese and Altaic without reaching a clear final conclusion. Forty-two years later, S. E. Martin (1966) provided 320 etymologies relating Japanese to Korean on the basis of regular sound correspondences, which

allowed him to reconstruct Proto-Korean–Japanese forms. R. A. Miller (1971), who established a set of sound correspondences to the Proto-Altaic phonemes reconstructed by Poppe (1960), clearly claimed Japanese to be one branch of the Altaic family. K. H. Menges (1975) took up a number of Miller's arguments and elaborated further on them. In his book on the Altaic problem and the origin of Japanese (1991), S. A. Starostin established sound correspondences between Japanese, Korean, and Altaic on the basis of numerous lexical comparisons of Turkic, Mongolic, Tungusic, Korean, and Japanese lexical items. J. Janhunen (1992, 1994), however, pointed out some problems with the Altaic affiliation of Japanese, which he considers premature. He takes Japanese and Ryukyuan to form a distinct family of its own and the Old Koguryō language, once spoken on the Korean peninsula, to be a close relative of Japanese.

Bibliography

- Benzing J (1953). *Einführung in das Studium der altaischen Philologie und der Turkologie*. Wiesbaden: Harrassowitz.
- Clauson G (1956). 'The case against the Altaic theory.' *Central Asiatic Journal* 2, 181–187.
- Clauson G (1962). *Turkish and Mongolian studies*. London: The Royal Asiatic Society.
- Comrie B (1981). *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Doerfer G (1963–1975). *Türkische und mongolische Elemente im Neupersischen, unter besonderer Berücksichtigung älterer neupersischer Geschichtsquellen, vor allem der Mongolen- und Timuridenzeit*. Wiesbaden: Steiner.
- Doerfer G (1966). 'Zur Verwandtschaft der altaischen Sprachen.' *Indogermanische Forschungen* 71, 81–123.
- Doerfer G (1976). 'Proto-Turkic: Reconstruction problems.' *Türk Dili Araştırmaları Yıllığı Belleten* 1975–1976, 1–50.
- Doerfer G (1985). 'The Mongol-Tungus connections.' *Language Research* 21, 135–144.
- Fokos-Fuchs D R (1962). *Rolle der Syntax in der Frage nach Sprachverwandtschaft (mit besonderer Rücksicht auf das Problem der ural-altaischen Sprachverwandtschaft)*. Wiesbaden: Harrassowitz.
- Janhunen J (1992). 'Das Japanische in vergleichender Sicht.' *Journal de la Société Finno-Ougrienne* 84, 145–161.
- Janhunen J (1994). 'Additional notes on Japanese and Altaic.' *Journal de la Société Finno-Ougrienne* 85, 236–240.
- Janhunen J (ed.) (2003). *The Mongolic languages*. London & New York: Routledge.
- Johanson L (1990). 'Zu den Grundfragen einer kritischen Altaistik.' *Wiener Zeitschrift für die Kunde des Morgenlandes* 80, 103–124.
- Johanson L (1999). 'Cognates and copies in Altaic verb derivation.' In Menges K H & Naumann N (eds.) *Language and literature – Japanese and the other Altaic languages*. Wiesbaden: Harrassowitz. 1–13.
- Martin S E (1966). 'Lexical evidence relating Korean to Japanese.' *Language* 42, 185–251.
- Menges K H (1975). *Japanisch und Altaisch*. Wiesbaden: Steiner.
- Miller R A (1971). *Japanese and the other Altaic languages*. Chicago & London: The University of Chicago Press.
- Miller R A (1996). *Languages and history. Japanese, Korean, and Altaic*. Bangkok: White Orchid & Oslo: The Institute for Comparative Research in Human Culture.
- Polivanov E D (1924). 'K rabote o muzykal'noj akcentuacii v japonskom jazyke (v svjazi s malajskim).' *Bjulleten' 1-go Sredne-Aziatskogo Gosudarstvennogo Universiteta* 4, 101–108.
- Poppe N (1960). *Vergleichende Grammatik der altaischen Sprachen 1. Vergleichende Lautlehre*. Wiesbaden: Harrassowitz.
- Poppe N (1965). *Introduction to Altaic linguistics*. Wiesbaden: Harrassowitz.
- Ramstedt G J (1924). 'A comparison of the Altaic languages with Japanese.' *Transactions of the Asiatic Society of Japan, 2nd ser.* 1, 41–54.
- Ramstedt G J (1939). *A Korean grammar*. Helsinki: Société Finno-Ougrienne.
- Ramstedt G J (1946–1947). 'The relation of the Altaic languages to other language groups.' *Finnisch-Ugrische Forschungen* 53, 15–26.
- Ramstedt G J (1949). *Studies in Korean etymology*. Helsinki: Société Finno-Ougrienne.
- Ramstedt G J (1952–1957). *Einführung in die altaische Sprachwissenschaft 1–2*. Helsinki: Société Finno-Ougrienne.
- Róna-Tas A (1986). *Language and history. Contributions to comparative Altaistics*. Szeged: University of Szeged.
- Róna-Tas A (1998). 'The reconstruction of Proto-Turcic and the genetic question.' In Johanson L & Csató É Á (eds.) *The Turkic languages*. London & New York: Routledge. 67–80.
- Starostin S A (1991). *Altajskaja problema i proisxoždenie japonskogo jazyka*. Moskva: Nauka.
- Tekin T (1986). 'Zetacism and sigmatism: main pillars of the Altaic theory.' *Central Asiatic Journal* 30, 141–160.

Amharic

D Appleyard, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introductory Remarks

Amharic (self-name *amaripṇa*) is the largest member of the South Ethiopic branch of Ethiopian Semitic languages. Amharic is spoken, according to the most recent estimate (1999), by around 17.4 million people as a first language and between 5 and 7 million more as a second language, making it the second largest Semitic language after Arabic, and the fourth largest language of sub-Saharan Africa after Swahili, Hausa, and Yoruba, although some estimates suggest that Oromo may have more speakers in total. Amharic is the main lingua franca of Ethiopia and is the constitutionally recognized working language of the country. As such it forms the language of instruction of public education at primary and secondary level, including from the third grade upwards in areas where it is not the first language. It is also the majority language of most urban-dwelling Ethiopians except where Tigrinya (Tigrigna) is the first language. The current status and wide distribution of Amharic are due especially to the amharization policies of previous Ethiopian governments in the 20th century. Until the change in language policy after the Ethiopian revolution of 1974, Amharic was the only Ethiopian language used in state education and the official media. The earliest records of Amharic date to the rise of the Amhara or Solomonid dynasty in the 14th century, and the spread of the language over an ever-increasing area of the Ethiopian highlands accompanied the expansion of the Christian kingdom up to modern times.

Modern Amharic shows some dialectal variation, though perhaps less than might be supposed for a language with such a wide distribution. This may in fact be due to the way in which the language has spread over the last 700 years, as part of a deliberate process of amharization, and it is notable to this extent that the dialect areas that are generally recognized are geographically defined within the regions where Amharic either originated or has been spoken the longest. The dialect of Shoa and, in particular, Addis Ababa has become the prestige dialect, forming a *de facto* standard. This is the form of Amharic that is used in the media as well as in the areas of administration and education.

Like all the modern Ethiopian Semitic languages, Amharic has been heavily influenced by the Cushitic languages alongside which it has developed, initially

the now minority Central Cushitic languages and then, as it spread, Highland East Cushitic and later Oromo. This influence can be seen not only in the lexicon, but also in syntax and typology. As the language of the ruling elite and thus the inheritors of Ethiopian Christian culture from Aksum, Amharic was also open to borrowing from Ge'ez, the classical or liturgical language of the Ethiopian Orthodox Church, which in more recent times has provided a rich source for the expansion of the Amharic lexicon to satisfy the need for technical, political, and other vocabulary.

Amharic is written in the Ethiopic syllabary, the script used for Ge'ez and developed in Ethiopia probably sometime during the 4th century C.E. out of the South Arabian consonantal alphabet. The Ethiopic syllabary, or *fidāl*, used for Amharic has 33 primary symbols, which indicate C + vowel /e/, each of which is further modified in some way to indicate C + one of the remaining six vowels: **ሀ** /be/, **ሁ** /bu/, **ሂ** /bi/, **ሃ** /ba/, **ሄ** /be/, **ህ** /bi/, **ሆ** /bo/, in the traditional sequence, giving 231 basic letters. Whilst some of the modifications are more or less regular across the whole system, others are not. For instance C + vowel /e/ is always marked by a loop attached to the bottom right-hand of the basic letter, but there are 16 different ways of marking C + vowel /i/. The whole structure is traditionally displayed in a grid with consonants on the vertical axis and vowels on the horizontal. The sixth column of the grid indicates both C + vowel /i/ and C without a following vowel: **ሐ** = both /bi/ and /b/. The contrast between C + /e/ and C + /a/ is mostly neutralized where C is a guttural **ኀ** or /h/: graphemes **ሀ** {hə} and **ሃ** {ha} are both /ha/. Whilst there are 33 base letters, these correspond to 27 consonant phonemes, as there is a certain amount of redundancy: for example, the letters **ሀ**, **ሁ**, **ሂ**, and **ሃ** all mark the consonant /h/; **ሐ** and **ሐ** mark a lack of consonantal onset, or **ኀ** depending on analysis. The labialized gutturals /k^w/, /g^w/, /k^w/, and /h^w/ are indicated by additional vowel symbols attached to the corresponding nonlabialized consonant signs: **ኸ** = /k^we/, **ኹ** = /k^wu/. In addition to these, a number of other consonant bases have a special symbol for C + /wa/: **ኸ** = /dʒe/, **ኹ** = /dʒwa/. There is lastly one other place where the Ethiopic syllabary does not correspond exactly to the phonemic structure of the language; consonantal length is phonemic in Amharic but is not marked at all in the script: thus /alɐ/ 'he said' and /allɐ/ 'there is' are both written **ሐ**, i.e., {ʔa} + {lɐ}. As an example of a piece of continuous text, consider the following, which is the last example cited in this article:

ይህ ነገር ብዙ ጊዜ ስለሚያስፈልግ አስከ

ምሽት ነው መሥሪያ ቤት የሚቆዩት።

{ji + hi ne + ge + ri bi + zu gi + ze

si + le + mi + ja + si + fe + li + gi ʔt + si + ke

mi + fe + ti ne + wi me + si + ri + ja be + ti

je + mi + k^we + ju + ti}

/jih neger bizu gize silemmijasfellig iske miffet dres

nɛw mɛsrija bɛt jɛmmik' ojjut/

'because this thing needs a lot of time, they'll stay

behind at work until evening'

Phonology

Amharic has a system of 30 consonant (see Table 1) and 7 vowel phonemes. Distinctive are the glottalized consonants, which have parallels in other languages of the Ethiopian language area. Also notable are the labialized gutturals /k^w/, /k^w/, /g^w/, and /h^w/; indeed, labialization of other consonants occurs, but only before the vowel /a/, and is contrastive as for instance in the nearly minimal pair /m^w at/ 'deceased' – /mɛtʃe/, /mɛtʃ/ 'when?' The addition of phonemic units such as /m^w/ would increase the number of consonant phonemes. Consonant length is also phonemic; only /h/ and the glottal stop, whose phonemic status in Amharic is debatable, do not have lengthened counterparts. The vowel system is distinguished by the presence of two central vowels, high /i/ and low-mid /e/, which together with low /a/ are the most frequent vowels in the language. Vowel length is not phonemic.

The vowels of Amharic are /i/, /i/, /u/, /e/, /o/, /e/, and /a/. The phonemic status of the vowel /i/ has been the matter of some discussion, and certainly its occurrence as a default epenthetic vowel in the application of syllable structure rules is predictable: the consonantal strings /s-n-t/, /m-l-kk-t/ being resolvable only as /sint/ 'how much?' and /milikkit/ 'sign,'

respectively. Contrast /d-n-g-l/, which surfaces predictably as /dingil/ 'virgin.' Indeed, the Ethiopic syllabary uses the same set of symbols for a consonant alone and a consonant + /i/. However, forms such as /jis'ifall/ 'he writes' rather than the predicted /*jis'fall/ indicate that /i/ does have phonemic status.

Ethiopianist convention occasionally employs different symbols from the IPA ones used here; thus, š = ʃ, ž = ʒ, č = tʃ, q = k', ɬ = t', ǧ = tʃ', ǧ = dʒ, ʃ = s', p = p', ñ = ɲ, y = j, ä = e, ə = i.

Syllable structure is [C]V[C][C], with no more than one consonant permitted in syllable onset position, and no more than two in syllable coda or, indeed, word medially and finally, with a lengthened consonant counting as two, as in the example of /milikkit/ above.

Accent in Amharic has been the subject of only a few studies, and its nature is still somewhat a matter of discussion. Generally, whilst Amharic accent is essentially a weak stress accent, it seems that word accent is subordinate to phrasal or sentence accent.

Morphology

Amharic has a complex inflectional morphology, particularly in the verbal system, employing not only prefixes and suffixes but also internal modification of the typical Semitic consonantal root-and-pattern type. In general, the morphology of Amharic has been less influenced by the Cushitic substratum than, for instance, syntax or the lexicon. The inflectional morphology of nouns, on the other hand, is relatively simple. Like other South Ethiopic languages, Amharic has mostly lost the heterogeneous system of noun plural formation by internal modification, the so-called broken plurals that are so common in North Ethiopic languages such as Ge'ez and Tigrinya, and in some other Semitic languages such as Arabic. Noun plurals in Amharic are for the most part formed by means of the suffix /-otʃtʃ/. Nouns also show two genders, though these are mostly manifest only in concord, chiefly between subject and verb predicate. Nouns further show definite marking by means of suffixes: masc. /-u/~/-/w/ and fem. /-wa/, which are in origin 3rd person pronominal suffixes: /bet-u/ is thus both 'the house' and 'his house.' Amharic does not have a true case system, adverbial functions being expressed variously by prepositions, or postpositions, or interestingly by a combination of the two: /kɛ-sɛwije-w gar/ 'with the man,' where /kɛ-/ and /gar/ together gloss 'with.' Of the primary relational case functions, the subject is unmarked, a definite direct object is usually marked by the clitic /-n/, which occurs after the marker of definiteness within the noun phrase, and the possessive or adjunct function is indicated by the bound preposition /je-/.

Table 1 The consonant phonemes of Amharic

| | bilabial | alveolar/ dental | palatal | velar | glottal |
|-------------------------------------|----------|---------------------|---|-------|----------------|
| Plosive/affricate | b p | d t | dʒ tʃ | g k | (ʔ) |
| Glottalized | p' | t' | tʃ' | k' | |
| plosive/ affricate/ fricative | | | s' | | |
| Labialized | | | g ^w k ^w k ^w | | h ^w |
| Fricative | f | z s | ʃ j | | h |
| Nasal | m | n | ɲ | | |
| Lateral | | l | | | |
| Approximant | w | r | j | | |

which is in form and origin identical to the adjunct or relative marker on verbs:

leba jə-gəbərə-w-n lam sərrək'-ə
thief of-farmer-DEF- cow steal.PAST-3MASC.
OBJ PAST

'a thief stole the farmer's cow'

abbat-e gomen-u-n b-atakilt bota zerra-[Ø]
father-my cabbage- in-vegetable place sow.PAST-
DEF-OBJ [3MASC.
PAST]

'my father sowed the cabbages in the garden'

The verb is inflected for voice or valency, tense-mood-aspect (TMA), and person. Negation is also marked within the inflected verb, as is to a large extent the distinction between main and subordinate verbs. In addition to the base stem, typically with active function, there are three fundamental voices or derived stems formed by prefixes: causative /a-/, passive-reflexive /tə-/, and factitive or (double) causative /as-/. There are other less productive formatives of more restricted occurrence, such as /astə-/, which also has a causative function, and /an-/ and /tən-/, with transitive-causative and stative-passive functions on verbs with expressive meaning (movement, sound, emotion, etc.). Internal changes in the various formations of TMA stems, however, combine with these prefix formatives and sometimes obscure them: /tə-sərrək'-ə/ 'it was stolen' but /ji-sərrək'-all/ 'it will be stolen,' where the imperfective or nonpast stem corresponding to /tə-sərrək'-/ is /-sərrək'-/. The occurrence of derived stem formatives is also to some extent lexical: /tə-k'əmmət'-ə/ 'he sat down' is active and does not contrast with a base stem /*k'əmmət'-/.

Other derived stem patterns involve internal modification such as a change of vocalization, or reduplication of syllables, often in combination with the prefixes described above: /a-nnəgaggər-u/ 'they engaged one another in conversation' from the basic /nəgaggər-u/ 'they spoke.'

TMA marking is done by internal changes in the verb stem together with variations in person marking. Most notable here is the use of one set of personal suffixes for the past in contrast to a quite different set of prefixes, or prefixes and suffixes combined, for the nonpast stem: /wəddək'-ətʃtʃ/ 'she fell' but /ti-wədk'-all-ətʃtʃ/ 'she falls, is falling,' /a-t-wədk'-imm/ 'she isn't falling,' /ti-wdək'/ 'let her fall,' /bi-t-wədk'/ 'if she falls,' etc., where the stems are past /wəddək'-/, nonpast /-wədk'-/, and jussive-imperative /-wdək'-/, and the person markers for the 3rd feminine are past /-ətʃtʃ/, nonpast and jussive /t[i]-/, and the other elements are variously /-all-/ main verb affirmative nonpast, /a- ... -[i]mm/ main verb negative nonpast, and /b[i]-/ 'if.'

In addition to subordinate verbs formed by prefixes such as the conditional formative above, Amharic also possesses an inflected all-purpose adverbial subordinate verb, called the gerundive in much of the literature, though the term 'converb' (CONVB), which is occasionally used, is a better label: /wədk'-a/ 'she having fallen,' but from /səmma-tʃtʃ/ 'she heard' /səmt-a/ 'she having heard.' The gerundive/converb is typically used in describing a sequence of events:

innantə izzih k'ərt-atʃtʃihu zimm
you.PL here remain.CONVB-2PL 'quiet'
bil-atʃtʃihu tə-k'əmmət'-u
say.CONVB-2PL sit.IMP-PL
'you, stay here and sit quietly' ('... being quiet')

t'əlat jəst-o tēmelləs-in
enemy flee.CONVB-3MASC return.PAST-1PL
'they enemy fled and so we returned'

The gerundive/converb in combination with the main verb marker (MVM) /-all/, etc., also forms the basis of a second past tense main verb form which generally indicates a recent past event or situation resulting from a past event: /alk'-o-all/>/alk'w-all/ 'it is finished.'

The formal distinction between main and subordinate verb forms is not carried through the whole TMA system. The past tense form, such as /wəddək'-ə/ 'he fell' occurs in both positions and has no MVM as such, whilst the simple nonpast form /ji-wədk'/ 'he falls, will fall' occurs only in subordinate position, either with an auxiliary as in /ji-wədk' nəbbər/ 'he was falling,' or more usually with a subordinating element: /jəmm-i-wədk'/ '(he) who falls,' /s-i-wədk'/ 'when he falls/fell.' When used in main verb position, it requires the partially inflecting MVM if affirmative: /ji-wədk'-all/ 'he falls,' /ti-wədk'-all-ətʃtʃ/ 'she falls,' or the main verb form of the negative marker if negative: /a-j-wədk'-imm/ 'he doesn't fall.'

In addition to the elements discussed so far, the verbal complex may also contain pronoun object markers. These are of two kinds, essentially direct object pronouns and pronominal object pronouns, which involve an element /-ll-/ or /-bb-/ clearly associated with the simple nominal prepositions /lə-/ 'to, for' and /bə-/ 'in, with':

ajt-ən-əw-all
see.CONVB-1PL-him-MVM
'we have seen him'

adrig-o-ll-ipp-all
do.CONVB-3MASC-for-me-MVM
'he has done [it] for me'

Syntax

Word order in Amharic is generally subject-object-verb (SOV), with subordinate clauses preceding the main clause. Noun phrases are also generally head final with modifiers, including relative clauses, preceding the noun. Whilst a large part of Amharic syntax is influenced by Cushitic language patterns and is in accord with the typology of verb-final languages, there are still structures such as prepositions alongside postpositions which betray the older 'classical' Semitic syntax. Like most languages of the Ethiopian language area, Amharic makes considerable use of focus marking, which is here expressed by a construction involving the copula, which 'highlights' the focused item, and the relative verb, the so-called cleft clause construction:

zəməd-otʃtʃ-wa n-atʃtʃəw bal
relative-PL-her COP-3PL husband
 jə-mərret'-u-ll-at
REL-choose.PAST-3PL-for-her
 'it is her relatives who have chosen a husband
 for her'

jih nəgər bizu gize
this thing much time
 silə-mm-ij-asfellig
because-REL-3MASC-need.NONPAST

iskə miʃʃət dirəs n-əw məsriya-bet
until evening until COP-3MASC work-place
 jəmm-i-k'ojj-u-t
REL.NONPAST-3(PL)-stay.NONPAST-PL-DEF
 'because this thing needs a lot of time, it's until
 evening that they'll stay behind at work'

Bibliography

- Appleyard D (1995). *Colloquial Amharic*. London/New York: Routledge.
- Bender M L & Fulass H (1978). *Amharic verb morphology*. East Lansing: African Studies Centre, Michigan State University.
- Cohen M ([1936] 1970). *Traité de langue amharique (Abysinie)*. Paris: Institut d'Ethnologie.
- Cowley R *et al.* (1976). 'The Amharic language.' In Bender M L, Bowen J D, Cooper R L & Ferguson C A (eds.) *Language in Ethiopia*. London: Oxford University Press.
- Hudson G (1997). 'Amharic and Argobba.' In Hetzron R (ed.) *The Semitic languages*. London/New York: Routledge.
- Kane T L (1990). *Amharic-English dictionary* (2 vols). Wiesbaden: Otto Harrassowitz.
- Leslau W (1995). *Reference grammar of Amharic*. Wiesbaden: Otto Harrassowitz.
- Ullendorff E (1965). *An Amharic chrestomathy*. Oxford: Oxford University Press.

Anatolian Languages

J G Macqueen, University of Bristol, Bristol, UK

© 1994 Elsevier Ltd. All rights reserved.

Strictly speaking the term 'Anatolian Languages' should refer to all the languages which are or have been in use in the region known as Anatolia (modern Turkey). In practice however the term is reserved for the Indo-European languages which were in use in that area in the second and first millennia BC (*see Indo-European Languages*).

The Anatolian Languages

For the second millennium, the most fully documented of these languages is Hittite (*see Hittite*), the main language of the extensive archives dated ca. 1650–1180 BC and preserved in cuneiform script on clay tablets at the site of Boğazköy (now Boğazkale) in central Anatolia. Less amply documented Anatolian languages from the same archives are Luwian and

Palaic, while a fourth language, written in a locally developed hieroglyphic script and preserved mainly on seal-impressions and on rock-monuments scattered over a wide area of Anatolia (there is evidence to suggest that it may also have been employed in documents written on wax thinly spread on wooden tablets) is rather clumsily known as Hieroglyphic Luwian or (less accurately) Hieroglyphic Hittite.

This language continued in use for inscriptions on stone in southeast Anatolia and north Syria well into the first millennium, while further west the local languages of Lycia and Lydia in the classical period, though written in scripts related to that of contemporary Greece, show clear signs that they too are members of the Anatolian group. Place names also provide evidence for the survival of Anatolian languages into the Roman period.

Phonology

In the area of phonology, a distinctive feature of the group is that Indo-European *o* is totally absent from

the vowel-system. But the most important distinguishing feature is the survival of at least some of the postulated Indo-European laryngeals which have been lost in all other groups. The nature and number of these laryngeals are still very much under discussion, but their appearance in the Anatolian languages offers strong support to the basic correctness of the theory first put forward by Saussure.

Morphology

The principal distinguishing characteristic of the group in the area of morphology is its lack of many features of the common Indo-European grammatical inventory. In the noun, for instance, the feminine gender is entirely absent, as is the dual number. Several parts of the plural paradigm are also lacking, although the singular retains a larger number of case forms. In the verbal system an even greater simplification has taken place, with only two moods (indicative and imperative) and only two tenses (present and preterite). Features such as reduplication and infixes -s-, elsewhere used in tense-formation, do exist, but they do not play any part in the Anatolian tense-system. There are two conjugations, known after the first present singular of each as the ‘*mi*-conjugation’ and the ‘*hi*-conjugation.’ Of these the former shows clear links with the Indo-European present-system, while the latter, though showing no ‘perfect’ characteristics in its use, seems to preserve in its endings elements of the Indo-European perfect. A medio-passive voice, with a similarly reduced mood- and tense-system, is also clearly attested.

Lexicon

Characteristic of the Anatolian lexicon is the extensive loss of original Indo-European vocabulary. Yet sufficient survives to indicate, as does the grammatical material, that the Anatolian languages, though subject throughout their history to a great deal of influence from non-Indo-European sources, still maintained their basic character as members of that family.

Particles

A lesser distinctive feature of the Anatolian languages is their liking for ‘chains’ of particles and enclitic pronouns placed at the beginning of a sentence or clause. Among these particles is one which serves the function of indicating indirect speech.

Division into Dialects

Study of the available texts has now made it possible to construct a dialect pattern of the Anatolian

languages. In the second millennium there is a clear distinction between northern (Hittite) and southern (Luwian) Anatolian. In phonology the main criterion is the treatment of Proto-Anatolian *e*, which in northern Anatolian with increasing closure moved towards *i*, while in the south it became more open and partially fused with *a*, thus obliterating the ablaut patterns which survived in the north. Among other distinctive features is the treatment of the voiceless dental before *i*. This is retained in the south, but affricated in the north; thus the 3pl ending is *-nti* in Luwian, but *-nzi* in Hittite. In the north too voiced dentals were assimilated before *i*, while in the south loss of voice was the rule (Hittite *siuni*- ‘god,’ *siwatt*- ‘day,’ as opposed to Luwian *Tiwat* ‘sun-god’).

In noun morphology the south shows a high proportion of *-i*-stems while the north retains a greater number of *-a*-stems; the north too shows a proliferation of *r/n*-stems in contrast to their disappearance in the south. The Indo-European nominative and accusative plural endings are retained in the north (Hittite *-es*, *us* < *ns*) but replaced in the south by Luwian *-nzi* and *-nza*, forms possibly of pronominal origin. The number of case forms, already reduced in Proto-Anatolian, is further reduced in the south, where in Luwian the genitive singular almost entirely disappears and is replaced by an adjectival suffix *-assi*-. In pronominal declension the south shows much more leveling with noun-forms than the north, while in the verbal system the principal southern distinction is the lack of the *-hi* conjugation present tense, although such forms as the Luwian first person singular preterite in *-ha* (not found in Hittite where the preterite is formed by the addition of secondary endings to the present stem) are ultimately related to the same source. Lesser distinctions are northern iterative *-sk*- as opposed to southern *-s(s)*-, and the retention in the south, but not in the north, of an archaic passive participle in *-mmi*-.

The features displayed by Palaic are mainly those of the northern subdivision, though some southern features (e.g., *e* > *a*, and the affrication of the voiceless dental before *i*) are clearly present. The language written in hieroglyphic script, on the other hand, is clearly southern in character, and is best described as East Luwian.

In the first millennium sources for North Anatolian are lacking, but East Luwian continues in use for several hundred years, showing a number of features which distinguish it from the Central Luwian of the previous period (e.g., nom pl in *-(a)i*, dat-loc pl in *-i*); and later still in western Anatolia, Lycian appears as a latter-day West Luwian language with its own local peculiarities (e.g., acc pl *-as*, dat-loc pl *-a* or *-e*, gen pl *-āi*; replacement of both Luwian *a* (< *e*) and Hittite

a by *e*). The position of Lydian is more difficult to establish. The apparent retention of *i* (< *e*), and the preponderance of *-a*-stems, for instance, point strongly towards the north, while features such as the disappearance of the genitive and its replacement by an adjectival suffix (in this case *-li-*) suggest a closer connection with the south.

Origins

Despite attempts to locate the 'homeland' of Indo-European within Anatolia itself, or immediately to the east of it, it is more generally accepted that the ancestor of the languages was introduced to the area from the north, more probably via the Balkans than via the Caucasus, and that the divisions described above took place in Anatolia during the third and early second millennia BC. The distinctive character of Anatolian, combining as it does extensive loss of original features (e.g., the feminine) with retention of other features which are extremely archaic (e.g., the laryngeals) makes it extremely likely that it diverged from the rest of the Indo-European continuum at an early stage, and was thus subject to a very long period of attrition from other languages with which it came into contact. There is however no need to postulate an earlier 'Indo-Hittite' from which the Anatolian languages on the one hand and

the Indo-European languages on the other are separately descended.

Bibliography

- Friedrich J (1960). *Hethitisches Elementarbuch I*. Heidelberg: Winter.
- Gusmani R (1964). *Lydisches Wörterbuch*. Heidelberg: Winter.
- Houwink Ten Cate P H J (1961). *The Luwian Population Groups of Lycia and Cilicia Aspera during the Hellenistic Period*. Leiden: Brill.
- Kammenhuber A *et al.* (1969). *Altkleinasiatische Sprachen*. Leiden: Brill.
- Keiler A R (1970). *A Phonological Study of the Indo-European Laryngeals*. The Hague: Mouton.
- Laroche E (1959). *Dictionnaire de la langue louvite*. Paris: Institut Français d'Archéologie d'Istanbul.
- Mallory J P (1989). *In Search of the Indo-Europeans*. London: Thames and Hudson.
- Puhvel J (1966). Dialectal aspects of the Anatolian branch of Indo-European. In Birnbaum H & Puhvel J (eds.) *Ancient Indo-European Dialects*. Berkeley, CA: University of California Press.
- Saussure F de (1879). *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*. Leipzig: Teubner.
- Steiner G (1990). The immigration of the first Indo-Europeans into Anatolia reconsidered. *Journal of Indo-European Studies* 18, 185–214.

Ancient Egyptian and Coptic

J D Ray, University of Cambridge,
Cambridge, UK

© 2006 Elsevier Ltd. All rights reserved.

The Ancient Egyptian language is first attested a little before 3000 B.C., when the earliest inscriptions in hieroglyphic make their appearance. Connected texts of some length are found from about 2700 B.C., and these develop into a considerable literature, which forms one of our major sources of information about the ancient Near East. The language survived the downfall of the Roman Empire and the transition to Christianity, and in its latest form, written in a modification of the Greek alphabet, it is known as Coptic. Coptic survived until well after 1000 A.D. Egyptian therefore has the longest attested history of any language, and this makes it uniquely important to linguistics. The language is a member of the Afroasiatic family (sometimes referred to as Hamito-Semitic), although its exact place within this family is

disputed. Many of the related languages were not written down until modern times, and several 'missing links' may never have been recorded at all. Egyptian shares the preference of most of this family for triconsonantal roots, from which whole families of words may be formed, normally through variations on the internal vowels and the use of some affixes. It may be this feature that encouraged the Egyptians to omit the vowels from their writing system. The language recognizes two genders, conventionally termed masculine and feminine; neuter meanings are expressed in the early stages of the language by the feminine, later by the masculine. It is possible that case endings, similar to those in some Semitic languages, existed at a very early stage of Egyptian, but they are not written and soon fell away. Traces may remain in the so-called construct state, where a direct genitive relationship is expressed by two nouns apparently in apposition. Grammatical function is marked by strict word order. A dual number is recognized alongside singular and plural.

The Egyptian verb has unique features. A stative tense, known in Coptic as the qualitative, seems to be inherited from an early stage of Afroasiatic, and has cognates in Akkadian (Egyptian). This tense expresses the result of a verbal action, and is often best rendered by an adjective or an adverbial phrase: ‘open, continuous, far away, already knowing,’ or the like. The narrative tense system, on the other hand, is peculiar to Egyptian, and appears to consist of various verbal nouns with possessive suffixes for subject (‘his hearing’ developing into ‘he hears’). Other forms include a possessive construction with parallels to modern perfects (‘hearing to him’ developing into ‘he has heard’), and an infixed series which expresses past, present, and future contingency. There is also a set of so-called active participles, which are really epithets or nouns of agent (‘a hearer’), and a sequence of relative tenses formed from passive participles (‘his heard one’ developing into ‘the one which he heard’). Participles and relative forms show two aspects, perfective and imperfective, depending on whether the action is envisaged as completed or not; there are also traces of a prospective, which has future or subjunctive force. Aspect also features in the narrative tenses, where prospective and probably circumstantial forms also occur. The language is VSO in narrative contexts, but stative constructions take the form SV. A remarkable feature is that four uses of the English verb ‘to be’ – existential, predicative, identifying, and partaking of a quality – are rendered by distinct constructions. On the other hand, there is no verb ‘to have,’ which is conveyed by periphrases such as ‘there is to me.’ A welcome omission is comparative inflection of adjectives: ‘she is better than I’ is expressed simply as ‘she is good against/in respect to/ me.’

This is the form taken by Egyptian in its classic period, Middle Egyptian, during the early second millennium B.C. This canonical stage was recognized by the Egyptians themselves, and was retained in formal inscriptions until the end of Pharaonic history. However, after about 1400 B.C., pressure from the spoken language, which was constantly changing, began increasingly to affect the written texts. The result is Late Egyptian, which took over many of the functions of its predecessor. Late Egyptian, which is the direct ancestor of Coptic, stands to Middle Egyptian rather as Italian does to Latin, although phonetic changes are often concealed by the continuity of the script. Word order is noticeably freer. The most obvious innovations are in the verb, where the old patterns are replaced by analytic expressions derived from obsolescent verbal forms. This process – which is strikingly similar to the development of modern English – leads to greater emphasis on time distinction and modal subtleties. The number of compound ‘tenses’ in

such a system is almost limitless, although one distinction present in the last phase of Late Egyptian – that between preterite and present perfect – is lost in Coptic. One unusual feature of Late Egyptian is the existence of a second series of tenses, which throw emphasis on an adverbial adjunct. These may have originated in the relative forms (‘what he heard (is) yesterday’ developing into ‘it was yesterday that he heard’). This system is foreshadowed in Middle Egyptian, although the details are not yet understood. The development of the verbal system makes Coptic appear an SVO language, although this is historically accidental. Coptic also dispenses with most adjectives, the passive voice, and most plurals, preferring stative paraphrases, using active third-person plural constructions, and marking the plural of nouns merely by the forms of the article, possessive adjective, or demonstrative. Late Egyptian contains many Semitic loanwords; Coptic, on the other hand, is almost as full of Greek words as modern English is of French or Latin.

Egyptian throughout its history deserves the epithet *lingua geometrica*, given to it in the 19th century, when the regularity and elegance of its constructions were first appreciated. The following examples may illustrate this. (Egyptian is conventionally transliterated into Romanized consonants.)

Middle Egyptian:

| | | | | |
|--------------|---------|---------|----------------|-----------|
| ḥʿ.n | h3b.n | wi | hm.f | r |
| arise+pa | send+pa | me | embodiment+his | to |
| K3š | r | sn-mw | sp, | ib.f |
| Cush | for | two+ord | occasion, | heart+his |
| 3w | im.i | r | bt | nbt |
| content+stat | in+me | against | thing | any+f/sg |

‘As a result his majesty sent me to Nubia for a second time, his heart being pleased with me more than anything.’

Late Egyptian:

| | | | | |
|--------------|--------------|-----------|-----------|---------------|
| wn.in | Pr-3 | hr | h3b.i | r |
| exist+pa | contingency | Pharaoh | upon | sending+me to |
| p3 | t3 | Nhs | n | p3 |
| the+m/sg | land | Nubian | in | the+m/sg |
| sp | mḥ-sn, | iw | h3ty.f | |
| time | filling+two, | situation | heart+his | |
| mtry | im.i | m | šsr | |
| content+stat | in+me | in | abundance | |

Coptic:

| | | | | | |
|--------------|--------|----------|------------|------------|------|
| ⲁⲩⲭⲟⲟⲥ | ⲛⲉⲓ | ⲟⲩⲁ | ⲛⲛⲉⲥⲛⲏⲩⲩⲩ | ⲭⲉ | ⲁⲛⲟⲕ |
| ⲛⲧⲁⲩⲩⲱⲁ | ⲁⲛ | ⲉⲛⲁⲩⲩ | ⲉⲛⲁⲩⲩⲉⲗⲁⲟⲥ | ⲉⲁⲓⲱⲛⲭ | |
| ⲭⲛⲛⲏⲛⲟⲩⲉ | ⲛⲁⲭⲟⲟⲩ | ⲧⲏⲣⲟⲩ | | | |
| afjoos | nci | oua | n ne | snēu | |
| pa+he+say+it | namely | one+m/sg | of the+pl | brother+pl | |

| | | | | | | |
|------------------|--------|------------------|------|----------|-----------------------|----------------|
| je | anok | n | ti | m | p | ša |
| saying | myself | not ¹ | 1/sg | in | the+m/sg | value |
| an | e | nau | e | p | | angelos, eai |
| not ² | to | look | at | the+m/sg | angel, situation+pa+I | |
| ōnh | hn | n | nobe | na | hoou | tērou |
| live | in | the+pl | sin | my+pl | day | entirety+their |

‘One of the brethren said, “For my part, I am not worthy to see the angel, having lived in sin all my days.”’

Bibliography

- Allen J P (2000). *Middle Egyptian*. Cambridge: Cambridge University Press.
 Callender J B (1975). *Middle Egyptian* (= Afroasiatic Dialects, vol. 2). Malibu, CA: Undena Publications.

- Gardiner (Sir) A H (1957). *Egyptian grammar* (3rd edn.). Oxford: Griffith Institute/Oxford University Press.
 Hodge C T (ed.) (1971). *Afroasiatic: a survey*. The Hague: Mouton.
 Junge F (2001). *Late Egyptian grammar*. Oxford: Griffith Institute/Oxford University Press.
 Lambdin T O (1983). *Introduction to Sahidic Coptic*. Macon, GA: Mercer University Press.
 Layton B (2000). *A Coptic grammar* (= Porta Linguarum Orientalium 20). Wiesbaden: Harrassowitz Verlag.
 Loprieno A (1995). *Ancient Egyptian. A linguistic introduction*. Cambridge: Cambridge University Press.
 Polotsky H J (1971). ‘Coptic.’ In Hodge C T (ed.). 67–79.
 Vergote J (1971). ‘Egyptian.’ In Hodge C T (ed.). 40–66.
 Vergote J (1973–1983). *Grammaire copte* (2 vols). Louvain: Edit Peeters.

Andean Languages

W F H Adelaar, Leiden University, Leiden, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

‘Andean languages’ is a cover term for the native indigenous languages spoken in the western part of South America, more precisely in the Andean mountain ranges and the adjacent Pacific coastal strip. Genealogically, the Andean languages do not constitute a unity. They comprise some language families, most of which have a limited geographical importance, as well as several linguistic isolates (languages without proven relatives or languages that have been left unclassified so far). An ‘Andean’ language family proposed by Greenberg (1987) covers only part of the Andean languages and has not been generally accepted. From a typological point of view, Andean languages are also highly diverse. Many Andean languages have become extinct and cannot be classified because of a lack of data.

From north to south, the following families and isolates are encountered. In the northern and eastern parts of the Colombian Andes, several languages belong to the Chibchan family, which extends further into Central America: Barí (Motilón; also in Venezuela), Chimila, Cuna (Kuna), Damana, Ika (Aruaco), Kogui (Cogui), and Tunebo (Uwa). The Muisca (Chibcha) and Duit languages, which have been extinct since the late 18th century, also belonged to the Chibchan family. Muisca, originally spoken in the surroundings of Bogotá, was a language of administration during the colonial period. Chibchan languages share a common

lexical base, but are highly diverse structurally. Some of them (Barí, Chimila) are tonal.

Chocoan, a small family comprising two languages, Waunana and Emberá, has its largest concentration in the Pacific regions of Colombia and Panama. It is one of the rare language groups in the Americas featuring ergative case. The Emberá, who occupy an expanding territory, are locally known under different names (Catío, Sambú, Saija, etc.).

Cariban, a large family with its center of gravity in the Amazonian region and in the Guyanas, is represented in the northeast of Colombia and in adjacent Venezuela by Opón-Carare (extinct), Yukpa (Motilón), and Japreria. Several extinct languages of the Magdalena valley received Cariban influence (Muzo, Colima, Panche, Pijao), although their exact classification remains undecided.

The Arawakan family, also one of the major Amazonian groupings, is represented on the Guajira peninsula, west of Lake Maracaibo, by two verb-initial languages, Guajiro (Wayuu) and Paraujano (Añú), a rarity for the Andean region. The Guajiro, with a population of about 300 000, are one of the fastest-growing indigenous groups in South America. Two small families, both extinct – Timote-Cuica and Jirajaran – were confined to the Venezuelan part of the Andes and its Caribbean foothills.

In the southern Andes of Colombia and adjacent Ecuador, the Barbacoan language family has five living members: Cayapa (Cha’palaachi, Chachi), Colorado (Tsafiki), Cuaiquer (Awa Pit), Guambiano, and Totoró. Several extinct languages (Cara, Pasto) may have belonged to this family, which extended

from the highlands to the Pacific Coast. In addition, several linguistic isolates are found in southern Colombia: Kamsá (Sibundoy), Páez (Nasa Yuwe), and the extinct Yurumanguí. On the coast of north-western Ecuador, the extinct Esmeraldeño (Atacame) language was also an isolate.

The central Andean region, which comprises the highlands and coast of Ecuador, Peru, and Bolivia, as well as northern Chile and northwestern Argentina, is dominated by two language families: Quechua(n) (see **Quechua**) and Aymaran (see **Aymara**). Both language groups are very similar from a phonological and structural point of view, and they share more than 20% of their lexicon. The Quechumaran hypothesis, which rests on these similarities, assumes that the two groups developed from a common source. However, nearly all the similarities can be explained by intensive contact (convergence), leaving the genealogical classification of both groups undecided. Quechua has about 8 000 000 speakers and is divided into numerous dialects with a limited degree of mutual intelligibility. Its territory extends from southern Colombia to northwestern Argentina with several interruptions. Aymaran comprises two, possibly three languages: Aymara (with over 2 000 000 speakers in Bolivia, Chile, and Peru), Jaqaru, and Cauqui (both in Peru). The typically agglutinating ('Altaic') structure based on suffixation of these languages has been considered characteristic for Andean languages, but the other languages in the region do not seem to share it in all respects. The Uru-Chipaya family, with one surviving language in Bolivia (Chipaya), has a different structure with some prefixation (along with suffixes) and extensive gender agreement.

The remaining languages of the central Andean region are all presumably extinct. They include (partly) documented languages, such as Atacameño (in northern Chile), Mochica (on the coast of northern Peru), and Puquina (in the border region of Bolivia and Peru). Some Puquina vocabulary (combined with Quechua morphology) survives in a professional language used by the Callahuaya herb doctors in Bolivia. Atacameño (Kunza) and Mochica are isolates, but Puquina may be distantly related to Arawakan. There is ample evidence of other, minimally documented languages: Panzaleo and the Puruhá-Cañar group in highland Ecuador, the Tallán-Sechura group (on the coast of northern Peru), Chacha and Culli (in the highlands of northern Peru), Quingnam (on the coast of central-northern Peru), Diaguita (in northwestern Argentina and in Chile), and Humahuaca (in northwestern Argentina). In addition, in Argentina the Lule or Tonocoté language (extinct but documented) presumably had its origin in the Chaco region.

In the southern Andes, Mapuche (Mapudungun; also known as Araucanian) is the native language with the largest distribution. Originally the dominant language of Chile, it is now confined to an area in southern Chile (Biobío, Malleco, Cautín, Arauco, etc.) and several locations in the Argentinian pampas and in Patagonia. Its number of speakers may be close to 500 000 (no reliable count is available). The closely related Huilliche (Tsesungun) language, originally spoken in Osorno Valdivia and on the isle of Chiloé, is nearly extinct. Mapuche is an agglutinating, suffixing language, as are Quechua and Aymara, but it differs from these languages in that it has practically no nominal morphology. By contrast, its verbal morphology is exceptionally rich. Some of its characteristics (interdental consonants, lack of case, noun incorporation) cause the Mapuche group to stand alone among the Andean languages. It has no known relatives. In the Argentinian region of Cuyo (Mendoza, San Juan), the unrelated Huarpean group (with the languages Allentiac and Millcayac) was spoken until the 17th century.

In the southern tip of Chile, the isolates known as Kawesqar (Qawasqar) or Alacaluf (in the archipelago west of the mainland) and Yahgan or Yamana (on the islands south of Tierra del Fuego) are both close to extinction. A third language, Chono (north of Kawesqar), has long been extinct. The Chon family, which comprises Ona or Selknam (on Tierra del Fuego), Tehuelche, Teushen, and Gününa Yajich (all in southern Argentina), is now only represented by Tehuelche, which is also nearly extinct.

An issue under debate is the affiliation of languages or language families situated on the eastern fringe of the Andes. From a genealogical viewpoint, this area is exceptionally diverse. Some of these languages share characteristics with Amazonian groups (e.g., 'Amazonian' classifiers, extensive prefixation, loose morphology, rich vowel systems, nasal harmony), whereas others are closer to Andean languages and seem to have had some relationship to the languages spoken in the highlands. Among the latter are Betoí and Cofán in Colombia (the latter also in Ecuador), the Jivaroan languages and the Candoshi group (in Ecuador and northern Peru), the Cahuapanan and Hibito-Cholón groups (in northern Peru), and a series of isolates on the Andean slopes of northern and eastern Bolivia (Leco, Mosetén, Movima, Yuracaré). Amuesha (Yanesha) found in Peru is an Arawakan language with a heavy Quechua admixture.

Because of massive language extinction, many Andean languages have disappeared during the last 500 years, leaving an incomplete picture of the original situation. It is not easy to link known languages to specific cultures established by archaeologists. Most

of the extinct languages were replaced by expanding local languages, such as Quechua, Aymara, and Mapuche, or by Spanish (now spoken by a majority of the population).

Bibliography

- Adelaar W F H & Muysken P C (2004). *The languages of the Andes*. Cambridge: Cambridge University Press.
- Albó X (1995). *Bolivia plurilingüe. Guía para planificadores y educadores*. La Paz, Bolivia: Unicef and CIPCA.
- Cerrón-Palomino R M (1987). *Lingüística quechua*. Cuzco: Centro Bartolomé de Las Casas.
- Cerrón-Palomino R M (2000). *Lingüística aimara*. Cuzco: Centro Bartolomé de Las Casas and PROEIB Andes.
- Chirinos Rivera A (2001). *Atlas lingüístico del Perú*. Cuzco and Lima: Centro Bartolomé de las Casas and Ministerio de la Educación.
- Constenla Umaña A (1991). *Las lenguas del area intermedia: introducción a su estudio areal*. San José, Costa Rica: Editorial de la Universidad de Costa Rica.

- González de Pérez M S & Rodríguez de Montes M L (eds.) (2000). *Lenguas indígenas de Colombia: una visión descriptiva*. Bogotá: Instituto Caro y Cuervo.
- Greenberg J H (1987). *Language in the Americas*. Palo Alto, CA: Stanford University Press.
- Hestermann D & Gusinde M (eds.) (1933). *Yamana-English: a dictionary of the speech of Tierra del Fuego*. Mödling: Anthropos Verlag (Original author: Th Bridges).
- Klein H E M & Stark L R (eds.) (1985). *South American Indian languages: retrospect and prospect*. Austin, TX: University of Texas Press.
- Mannheim B (1991). *The language of the Inka since the European invasion*. Austin, TX: University of Texas Press.
- Mosonyi E E & Mosonyi J C (2000). *Manual de lenguas indígenas de Venezuela*. Caracas: Fundación Bigott.
- Salas A (1992). *El mapuche o araucano. Fonología, gramática y antología de cuentos*. Madrid: Editorial MAPFRE.
- Torero Fernández de Córdova A A (2002). *Idiomas de los Andes: lingüística e historia*. Lima: Instituto Francés de Estudios Andinos and Editorial Horizonte.

Arabic

S Procházka, University of Vienna, Vienna, Austria

© 2006 Elsevier Ltd. All rights reserved.

Arabic is the official language of 21 countries in the Middle East and North Africa, from Oman in the east to Mauritania in the west. This includes Israel, where Arabic is, after Hebrew, the second official language. Significant Arab minorities exist in Iran, Turkey, Chad, and Nigeria, as well as in western Europe and the Americas. With approximately 280 million native speakers, Arabic is by far the largest living representative of the Semitic language family. Because it is the language of the Koran and thus the liturgical language of Islam, Arabic also plays an important role for more than 1 billion Muslims worldwide.

History of the Language

Arabic is an offshoot of the Semitic branch of the Afro-Asiatic languages. According to the traditional classification of Semitic, Arabic is part of its southern subdivision and grouped with Ethiopic and South Arabian (by stressing the common $p > f$ shift and the internal plurals). In the 1970s, Hetzron proposed placing Arabic with Aramaic and Canaanite in a 'Central Semitic' group (stressing the imperfect pattern and the t as a marker for the first- and second-person singular perfect). The problem of the affiliation

of Arabic within the Semitic languages continues to be discussed (see Faber, 1997).

Although people labeled Arabs are attested as early as the 9th century B.C.E. in Assyrian sources, the history and development of their language before the emergence of Islam, 1.5 millennia later, is largely unknown. Doubtless Arabic originated in the central and northern parts of the Arabian peninsula, later spreading northward to the edges of the Fertile Crescent. The first evidence of a language akin to Arabic are the so-called Ancient North Arabian inscriptions (5th century B.C.E. to approx. 4th century C.E.): these consist of thousands of short, and therefore linguistically scarcely informative, graffiti in a script derived from the South Arabian writing system and found mainly in western Arabia and southern Syria. There are traces of Arabic in the Aramaic inscriptions of the Nabateans and Palmyrenes – both certainly Arab people. Textual evidence of pre-Islamic Arabic is also found in a handful of inscriptions in early Arabic script from the 2nd to 6th centuries C.E.

Our richest source of pre-Islamic Arabic is a large corpus of orally transmitted poetry from the 6th and 7th centuries C.E., later compiled by Arab philologists. The language of these poems and, although not exactly identical to theirs, that of the Koran (proclaimed by Muhammad between circa 610 and 632) is usually termed 'Old Arabic.' These texts, although a kind of poetic koinè, contain phonetical,

morphological, and lexical inconsistencies that reflect the actual dialectal differences between the spoken tribal vernaculars of the era (on these, see Rabin, 1951).

The expansion of Arab territory during the Islamic conquests (7th–8th centuries) made Arabic the language of communication, administration, and liturgy for an empire that stretched from central Asia to the Atlantic. The form of Arabic described, systematized, and canonized by the Arab grammarians and lexicographers between the 8th and 10th centuries is called Classical Arabic (CA). It remains the only universally accepted standard of the language. During the Golden Age of the Abbasid caliphate (9th–10th centuries) CA became the linguistic vehicle of a highly developed civilization that brought forth a rich literature, including *belles-lettres* and religious and scientific works. The hegemony of Arabic during the Middle Ages, and its prestige as the ‘sacred’ language in which the holy book of the Koran had been revealed to humankind, have influenced the languages of all Muslim people, written and unwritten. Thus, the lexicon of languages such as Persian (Western Farsi), Urdu, Turkish, or Swahili include numerous CA words. In many Muslim countries, Arabic has continued to be the language of religious treatises, and the teaching of it forms part of school curricula.

The Present Situation

Modern Standard Arabic

During Ottoman rule over most parts of the Arab world (from the 16th century onward), Arabic stagnated linguistically and literarily. Thus, in the early 19th century, when Arab intellectuals began to ‘discover the West’ and to translate European works into Arabic, they soon recognized its lexical shortcomings. This was the starting point of Modern Standard Arabic (MSA). MSA is practically identical in phonology, morphology, and syntax to CA, but it exhibits major differences from it in lexicon, phraseology, and style. After World War I, the modernization of Arabic continued in the language academies of Damascus, Cairo, and other capitals, which coined and still are coining thousands of neologisms. But not all the problems have been solved and, particularly in technical and scientific terminology, Arabic has not yet reached the standard of European languages. Competition among the academies frequently resulted in several terms for one and the same thing, and many academic neologisms have not been accepted by the speech community, which often prefers a loanword from English or French. In the standard language, loans play a remarkably minor role, but

the phraseology and style of MSA is deeply influenced by English (and in the Maghreb by French), above all in the language of the media. Thus, it is justified to call MSA a register of Arabic clearly differentiated from the classical language. The importance of MSA is that, as the only accepted medium of written and formal oral communication, it constitutes the tie that linguistically binds the Arab world together. However, MSA has to be learned in school because the native tongue of every Arabic speaker remains his or her local dialect as used in everyday life by all social strata. Therefore, MSA is almost completely limited to written use and to highly formal speech (news, official speeches, and academic discourse). Actually, this diglossic situation has been inherent in Arabic for at least the past millennium. The two linguistic layers are, of course, in a state of permanent mutual influence, and between the extremes of ‘pure standard’ and ‘plain colloquial’ Arabic are levels such as ‘educated colloquial.’ During the past decades, active and, especially, passive knowledge of MSA has significantly increased because of better education and the media. This trend was recently reinforced by the establishment of pan-Arabic satellite channels, which enjoy great popularity. Thus, even if MSA remains restricted to the domain of written and formal speech, a continually growing portion of the speech community will be able to participate in it.

Arabic Dialects

The various dialects belong to a language type called ‘New Arabic,’ whereas both CA and (in spite of its label ‘modern’) present-day MSA are ‘Old Arabic.’ The term ‘Middle Arabic’ does not denote, as we might assume, an intermediate chronological stage but a form of written Arabic exhibiting deviations from the standard norm due to the influence of ‘New Arabic,’ (i.e., the dialects; see Veersteegh, 1997: 114–129).

Although there are numerous typological differences, it is widely accepted, especially among Arabic speakers themselves, that the distinction between Old and New Arabic is the presence or absence of the case and mood endings (in Arabic, *iʿrāb*). The question of when and how the transformation from the old to the new type of Arabic happened is one of the most intriguing and discussed issues of Arabic studies (good summaries are Holes, 1995: 7–14; Versteegh, 1997: 93–113). There are indications from inscriptions that in the speech of the Nabateans the case system may have broken down as early as the 1st century C.E. If this is true, the new type of Arabic would have been spread along the trade routes of northern and western Arabia before the rise of Islam. Nevertheless, it seems very likely that in the time of

Muhammad the structure of everyday Arabic was not identical, but quite close, to the language of the poetry and the Koran. Only the social and political turmoils during and after the conquests resulted in a rapid shift to New Arabic. It should be emphasized, however, that Arabic developed along a line of internal linguistic trends common to all modern Semitic languages and clearly traceable before that time. The argument, often urged by the Arabs themselves, that these changes were mainly caused by so many non-Arabs using Arabic must be rejected.

The new type of Arabic spread among the urban centers of the Fertile Crescent and Egypt (the countryside had not yet been Arabicized) in the aftermath of the conquests. The language of the Bedouins, however, was not, or was only slightly, affected by these changes until approximately 2 centuries later. Ferguson (1959) explained the relative homogeneity of the urban dialects by the existence of a single koinè in the 7th–8th centuries. Although this theory is not tenable in its entirety, it was the starting point of a fruitful scientific discussion. From the present point of view, it seems very likely that the resemblances among the urban dialects are the consequence of continuous convergence and the mutual leveling of several regional koinai (see the summary in Miller, 1986).

The greatest typological differences are found between the sedentary (urban and rural) dialects and the Bedouin dialects. Thus, the speech of a sedentary Bedouin living in the outskirts of Tunis, for example, typically is closer to that of a Bedouin of Mauritania living 2000 miles away than it is to the speech of his neighbors speaking the dialect of the city of Tunis. Another sharp division separates the North African or Maghrebi dialects (including Maltese) west of Egypt from those to the east. The eastern dialects themselves can be divided into four large groups: (1) Arabian Peninsula, (2) Mesopotamia, (3) Syria and Palestine, and (4) Egypt, Sudan, and Chad (see Fischer and Jastrow, 1980). Audio files of a great number of dialects are available on the Semitic Sound Archive website of the University of Heidelberg.

Structure of Arabic

Phonology

The Arabic vowel system consists of three vowels /a, i, u/, with a phonemic contrast of short and long, for example, [mudi:runa:] ‘our director’ versus [mudi:ru:na:] ‘our directors’. In contrast to this relatively small number of vowels, Arabic possesses 28 consonant phonemes (see Table 1), also with a phonemic short-long contrast, for example, [hama:m] ‘pigeons’ versus [ham:a:m] ‘bath’ (as is usual, isolated Arabic nouns are cited without their case endings). The characteristic sound of Arabic is created mainly by a couple of consonants articulated in the velar and postvelar regions of the vocal tract and by the four velarized (also called ‘emphatic’) consonants that also have a lowering effect on adjacent vowels.

The realization of the consonant phonemes in MSA reflects almost completely the situation of Old Arabic. Exceptions are *j* [dʒ] (ج), which was most probably pronounced [ʝ], and the somewhat problematic sound *ḍ* [ɢ] (ض). There is an ongoing discussion on the original pronunciation of this consonant, which was so characteristic that Arabic was even called ‘the language of the letter *ḍād*’ (*luḡhat aḍ-ḍād*). Most likely it was either a velarized lateral fricative [ɬ] or a lateralized variety of *ḍ* [ɢ^l] (the latter perhaps reflected in such Spanish loans from Arabic as *alcalde* < ‘*al-qāḍī*’ ‘the judge’).

The present-day standard pronunciation of the consonants shows no regional variations other than the sound *ḍ* [ɢ] (ظ), which in many countries (e.g., Syria and Egypt) is pronounced *z* [z].

Except in religious utterances (i.e., the recitation of the Koran), other alterations are widely accepted, which make it quite easy to recognize the country of a given news broadcast. The most striking among these is the replacement of [dʒ] by [g] in Egypt or by [ʒ] in the Levant and large parts of North Africa.

The syllabic structure of CA is restricted to three types: CV, CV:, and CVC (under certain conditions also Ca:C/CayC). However, in MSA final short

Table 1 The consonants of standard Arabic

| | Bilabial | Labiodental | Dental/Alveolar/Postalveolar | Palatal | Velar | Uvular | Pharyngeal | Glottal |
|---------------------|----------|-------------|---------------------------------------|---------|-------|--------|------------|---------|
| Plosive | b | | t d ḏ | | k | q | | ʔ |
| Nasal | m | | n | | | | | |
| Trill | | | r | | | | | |
| Fricative | | f | θ ð ṣ s ʂ z (ʒ) ^a ʃ dʒ | | x ɣ | | ħ ʕ | h |
| Approximant | | v | | j | | | | |
| Lateral approximant | | | l | | | | | |

^aThe sound /ʒ/ is used in some countries (e.g., Egypt, Syria, and Lebanon) instead of ṣ.

vowels are often omitted, so CV:C and CVCC are also found. An Arabic word cannot begin with a vowel, and two vowels must be separated by no fewer than one consonant but by no more than two consonants.

Phonology of the Dialects Leaving aside the lexicon, the greatest difference among Arabic dialects is in phonology. The following summary provides only a general overview. In all modern dialects (with negligible exceptions in Yemen) the non-initial [ʔ] is lost and the two sounds ڤ (ظ) and ڤ (ض) are mingled into one sound [ð]. The consonants that most frequently exhibit changes compared to CA are (1) the three interdental fricatives [ð], [θ], and [ð̤] that, in the majority of the sedentary dialects, have been shifted to corresponding postdental stops (i.e., [d], [t], [d̤]); (2) the affricate [dʒ], which is pronounced [ʒ] in Central Arabia and the Sudan, [ʒ] in large parts of North Africa and the Levant, [g] in Lower Egypt, and [j] along the Arab Gulf; and (3) the reflexes of CA *q* (usually indicates whether a dialect is of the Bedouin or the sedentary type), which in Bedouin dialects has a voiced pronunciation ([g], [dz], [dʒ]) but in sedentary dialects is usually unvoiced ([q] or, as a typical urban phenomenon, [ʔ]).

Excluding the few that have been lengthened, all final short vowels of CA have been lost. There is also an almost universal tendency toward eliding unstressed short vowels (especially [i] and [u]) in open syllables (e.g., Cairo (Egyptian Spoken Arabic): [ʔjirib] ‘he drank’ versus [ʔjirbu] ‘they drank’). Many sedentary dialects exhibit a reduction of the inventory of short vowels from three to two (either a/ə or u/ə), whereas the majority of both Bedouin and sedentary dialects have developed a system of five long vowels [a:, e:, i:, o:, u:] as a result of the monophthongization of [ai] > [e:] and [au] > [o:].

Morphology

Derivational Morphology In all layers of Arabic, the bulk of the vocabulary is built on the principle of root and pattern. To express certain semantic terms (i.e., words), a purely consonantal root carrying the basic semantic information is combined with a limited set of patterns using a fixed sequence of consonants, vowels, and optional prefixes and suffixes. Most of the roots consist of three consonants called radicals. Those with four consonants are by no means rare, but are often merely extensions of triconsonantal roots. A few words of the most elementary vocabulary have only two radicals, for example, *ʾab* ‘father’, *yad* ‘hand’, and *mā* ‘water’. Such words, and the numerous instances of triconsonantal roots with two

common radicals expressing similar semantic concepts, have fueled speculations that the original system was built on a biconsonantal root system.

Many patterns are semantically and morphologically ambiguous; that is, one and the same pattern can serve for different semantic concepts and can be used for both verbs and nouns and for both singular and plural. Nevertheless, there are also patterns that are used exclusively for verbs or for certain semantic or morphological classes.

- CuCayC is the pattern of diminutives, for example, *kuwayt* ‘small fortress’.
- maCCaC/-a is used for nouns of place, for example, *maktab* ‘office’, *maktaba* ‘library’ (root *k-t-b* ‘writing’).
- miCCaC/miCCāC is used for instruments, for example, *miṣʿad* ‘elevator’ (root ṣ-ʿ-d ‘ascending’), *miṭṭāḥ* ‘key’ (root *f-t-h* ‘opening’).
- CaCCāC denotes professions, for example, *jazzār* ‘butcher’ (root *j-z-r* ‘slaughtering’).
- CaCCāCa is used for professions of females and instruments, for example, *ghassāla* ‘washer-woman, washing machine’ (root *gh-s-l* ‘washing’), *barrāda* ‘refrigerator’ (root *b-r-d* ‘cold’).

As can be seen from *miṣʿad* and *barrāda*, the system of derivation is widely used for the creation of neologisms. Although noun patterns are quite numerous (approximately 90 in CA) and are mostly not clearly related to semantic classes, the derivation of verbs is practically limited to 10 stems for triconsonantal roots and two for quadriconsonantal roots. Each stem has a set of five patterns reserved for the perfect and imperfect base, for the active and passive participle, and for a verbal noun (also called infinitive, lexicalized, i.e., not predictable, in stem I). As is shown in Table 2, certain functions can generally be attributed to each verb stem, although in detail the situation is highly complex (see the overview in Cuvalay-Haak, 1997: 95–108).

The principle is exemplified by the root *q-ṭ-ʿ* ‘cutting’.

- I: *qaṭaʿ-a* ‘to cut (in two)’.
- II: *qaṭṭaʿ-a* ‘to cut into pieces’.
- III: *qāṭaʿ-a* ‘to dissociate’.
- IV: *ʾaqṭaʿ-a* ‘to make cut’.
- V: *taqaṭṭaʿ-a* ‘to be cut off’.
- VI: *taqāṭaʿ-a* ‘to break off mutual relations’.
- VII: *ʾinqaṭaʿ-a* ‘to be cut off’.
- VIII: *ʾiqṭaṭaʿ-a* ‘to take a part’.

Note that no root is combined with all 10 stems.

The root-pattern system of derivation is responsible for the remarkable uniformity of the Arabic lexicon. Only a very few types of roots, above all those

Table 2 Stems of triconsonantal verbs in Standard Arabic

| Stem | Perfect | Imperfect | Verbal noun | Active participle ^a | General functions | Frequency/MSA ^b |
|------|------------|--------------|--------------------------|--------------------------------|----------------------|----------------------------|
| I | CaCVC- | ya-CCVC- | CVCC/CVCVC ^c | CāCiC- | basic | 40.07% |
| II | CaCCaC- | yu-CaCCiC- | taCCiC- | muCaCCiC- | causative/intensive | 14.28% |
| III | CāCaC- | yu-CāCiC- | muCāCaC-at- ^d | muCāC:C | conative and others | 5.14% |
| IV | 'aCCaC- | yu-CCiC- | 'iCCāC:C | muCCiC- | causative/factitive | 10.56% |
| V | taCaCCaC- | ya-taCaCCaC- | taCaCCuC- | mutaCaCCiC- | reflexive/passive | 10.80% |
| VI | taCāCaC- | ya-taCāCaC- | taCāCuC- | mutaCāCiC- | reciprocal | 4.44% |
| VII | 'inCaCaC- | ya-nCaCiC- | 'inCiCāC- | munCaCiC- | intransitive/passive | 2.93% |
| VIII | 'iCtaCaC- | ya-CtaCiC- | 'iCtiCāC- | muCtaCiC- | reflexive | 6.94% |
| IX | 'iCCaCC- | ya-CCaCC- | 'iCCiCāC- | muCCaCC- | colors ^e | 0.19% |
| X | 'istaCCaC- | ya-staCCiC- | 'istiCCāC- | mustaCCiC- | reflexive and others | 4.67% |

^aThe passive participle has an *a* instead of *i* in the last syllable, except in stem I, where the pattern maCCūC-is used.

^bRelative frequency of the stems in a modern dictionary; from Cuvalay-Haak (1997: 88).

^cBoth occur also with the suffix-at; there are numerous other patterns, in CA approximately 40.

^dAnd CiCāC-.

^eFor instance 'ihmarr-a 'to be red, to blush'.

containing the two weak consonants *w* and *y*, cause changes in most patterns; but, because even these follow certain rules, Arabic morphology is almost completely free of irregularities.

Noun Inflection The class of nouns comprises substantives, adjectives, and numerals; the categories gender, number, definiteness/indefiniteness, and case are differentiated. Arabic has two genders, masculine and feminine, the latter marked usually by the suffix-*a(t)* and in some noun patterns by -*ā'*/*-ā*. Among the unmarked feminines are nouns denoting beings of the female sex (e.g., 'umm 'mother'), paired parts of the body (e.g., 'ayn 'eye'), and some basic concepts of nature (e.g., 'arḍ 'earth', shams 'sun', and nār 'fire').

The number system is trifold: singular (unmarked), dual (suffix -*āni*), and plural. The plural is formed either by suffixation (MASC PL -*ūna*; FEM PL -*āt*) or more frequently by a complete restructuring of the word (thus the term internal or 'broken' plural), for example, bayt 'house', buyūt 'houses'; kitāb 'book', kutub 'books'; miftāḥ 'key', mafātīḥ 'keys'. A number of patterns (especially those containing three vowels) are restricted to plurals, but many others are used for both numbers; the pattern CiCāC, for instance, is singular in kitāb 'book', but plural in jibāl 'mountains' (for broken plurals see, Murtonen, 1964). Indefiniteness is usually expressed by a final -*n*, for example, bayt-u-n 'a house'; definiteness is usually expressed by the proclitic article 'al- (assimilated to dentals, sibilants, *n*, and *r*), by a pronominal suffix, or by a following genitive, for example, 'al-bayt-u 'the house', bayt-u-nā 'our house', bayt-u ḥasan-i-n 'Hasan's house'.

Arabic has the three cases, nominative, genitive, and accusative, which are differentiated in the singular and in broken plurals by declensions marked by

the final vowel, for example, NOM 'al-bint-u, GEN 'al-bint-i, ACC 'al-bint-a 'the girl'. The dual and external plural have common forms for the genitive and accusative (DUAL -*ayni*, MASC PL -*ina*, FEM PL -*āt-i-n*), a feature that is shared by a second (called 'diptote') type of declension (NOM -*u*, GEN/ACC -*a*) used primarily in female or foreign personal names and in certain plural patterns (in indefinite status).

Pronouns In pronouns, and hence in verbal inflection, Arabic distinguishes between masculine and feminine in all but the first person and the dual (see Table 3). Independent personal pronouns exist only in the nominative; for the other cases, suffixed forms are used, for example, 'anti mariḍ-at-u-n 'you (FEM SING) are ill', bayt-u-ki 'your (FEM SING) house', qabbala-ki 'he kissed you (FEM SING)'.

The relative pronouns and the two sets of demonstrative pronouns (for near and far deixis) also differentiate gender and number.

Verb Inflection Arabic has a twofold system for the inflection of finite verbs: a suffix-based conjugation, traditionally called 'perfect', and a prefix-based conjugation, called 'imperfect'. For both of these bases, a second set of vowel patterns exists to form a passive voice, for example, in stem I ḍarab-a 'he hit' versus ḍurib-a 'he was hit'; ya-ḍrib-u 'he hits' versus yu-ḍrab-u 'he is hit'. Usually the passive is used when the agent of a sentence is not mentioned or to express impersonality, for example, ya-dkḥul-u 'he enters' versus yu-dkḥal-u 'one enters'.

The imperfect has four moods, morphologically marked by different suffixes (the examples in parentheses are the forms of 'to write' in third-person singular masculine): indicative (ya-ktub-u), subjunctive (ya-ktub-a), jussive (ya-ktub-Ø), and

Table 3 Personal pronouns

| | Singular | | Dual | | Plural | |
|--------|-------------|----------------------|-------------|-------------|-------------|---------------|
| | Independent | Suffixed | Independent | Suffixed | Independent | Suffixed |
| 1 | 'anā | -nī -ī ^a | | | nahnu | -nā |
| 2 MASC | 'anta | -ka | 'antumā | -kumā | 'antum | -kum |
| 3 FEM | 'anti | -ki | 'antumā | -kumā | 'antunna | -kunna |
| 3 MASC | huwa | -hū/-hī | humā | -humā/-himā | hum | -hum/-him |
| 3 FEM | hiya | -hā | humā | -humā/-himā | hunna | -hunna/-hinna |

^a-nī is used with verbs, -ī with nouns and prepositions.

the so-called energetic (*ya-ktub-anna*), which is used in CA to express very strong assertions but is almost obsolete in MSA. The imperative is basically a subset of the jussive without prefixes. The verb conjugation expresses person, gender, and number. The system is, except for an additional dual for third-person feminine, analogous to the pronouns given in Table 3.

Morphology of the Dialects Generally speaking, no radical structural changes appear in the morphology of the dialects as compared to CA. Morphological derivation by applying the principle of root and pattern has been slightly simplified (there are fewer patterns compared to CA), but has remained productive. The most striking morphological difference between CA/MSA and all dialects is the lack of a case system. The indefinite marker *-n* has not survived except in some Bedouin dialects where it is found in a few syntactical positions such as attribution (e.g., North Syrian Bedouin: *bēt-in chibīr* 'a big house'). Some dialects (e.g., Iraqi), however, have developed an indefinite article.

All dialects lack dual forms of the pronouns and verbs, and most sedentary dialects have given up gender distinctions in the plural and those in North Africa no longer have gender distinctions in the second-person singular, as well. Together with nouns, the category dual is fully productive in the east, but in the sedentary dialects west of Egypt the dual is usually expressed by the numeral 'two' followed by a noun in the plural.

For the verbs, the perfect conjugation has not changed significantly. In the imperfect, however, the category mood is not expressed by internal inflection (a result of the loss of final short vowels) but, instead, where not completely obsolete, by modifiers prefixed to the verb. For example, in Damascus *b-yashrab* 'he drinks' roughly corresponds to the indicative and *yashrab* to the subjunctive/jussive. The formation of an internal passive voice seems to be limited to a few Bedouin dialects. In other dialects, certain verbal

stems (especially VII and VIII) are used to express passive voice, for example, Damascus: *ḥabas* 'he imprisoned' (stem I) versus *nḥabas* 'he was imprisoned' (stem VII).

Syntax

Tense and Aspect The verbal system of Arabic can be described as a combination of aspect and time reference. The suffix conjugation (called 'perfect') serves for the past and for the perfective (completed/factual) aspect, and the prefix conjugation serves for the nonpast (present/future) and for the imperfective (noncompleted/ongoing) aspect, including habituality, continuousness, and progressivity. An exception is the combination of the negation *lam* and the jussive mood, which indicates the negation of the perfect (e.g., *lam ya-ktub* 'he has not written').

The Arabic tense system is to a high degree a relative one. In main clauses, the temporal reference point is usually the moment of speaking, whereas in complement clauses the time has to be derived by reference from the main verb. Verbs in the perfect are also used in conditional clauses, in wishes and curses, and for assertions of factuality:

Allāh-u 'azz-a wa-jall-a
 God-NOM was.mighty-PERF and-was.sublime-PERF
 'God, he is mighty and sublime'

Participles do not mark any particular time reference, but frequently serve for a resultant aspect; that is, they describe an action that bears relevance to the moment of speaking.

Word Order The basic neutral word order of Arabic is VSO, but thematization of the subject is achieved by SVO. The latter therefore is not possible if the subject is indefinite, in which case sometimes also VOS appears.

The foreground/background distinction also influences word order. Generally VS is used for foreground information and events, and SV for background information and descriptions.

An adjectival attribute follows its head noun and agrees with it in case, in definiteness, and – with restrictions – in gender and number:

| | |
|---|---|
| bayt-u-n <i>house.MASC.SING-NOM-</i> INDEF | kabīr-u-n <i>big.MASC.SING-NOM-</i> INDEF |
| ‘a big house’ | |
| fi l-qal‘-at-i <i>in the.DEF-fortress.SING-FEM-</i> GEN | l-kabīr-at-i <i>the.DEF-big.SING-FEM-</i> GEN |
| ‘in the big fortress’ | |

Nominal annexations are in the genitive case and follow the head noun, which is morphologically determined (i.e., in the so-called *status constructus*). Indefinite/definite is therefore indicated solely by the noun annexed to it, for example, *bāb-u bayt-in* ‘a door of a house’ versus *bāb-u l-bayt-i* ‘the door of the house’. Although the number of annexations is theoretically unrestricted, there can be only one head noun. In phrases such as ‘the director and the teachers of the school’ the second head noun follows the genitive and takes a suffix referring to it:

| | | |
|---|--|--|
| mudīr-u <i>director.MASC-</i> NOM | l-madras-at-i <i>the-school-</i> FEM-GEN | wa-mu‘allim-ū-hā <i>and-teachers-</i> NOM-PL-her.FEM |
| ‘the director and the teachers of the school’ | | |

Under the influence of European languages, this rule is frequently ignored in MSA.

Agreement Strict agreement in gender and number exists only in the singular. Nouns in the plural agree with feminine singular unless they denote human beings.

| | |
|---|---|
| kutub-u-n <i>books.MASC.PL-NOM-</i> INDEF | qayyim-at-u-n <i>precious-FEM.SING-NOM-</i> INDEF |
| ‘precious books’ | |
| rijāl-u-n <i>men.MASC.PL-NOM-</i> INDEF | kirām-u-n <i>generous.MASC.PL-NOM-</i> INDEF |
| ‘generous men’ | |
| ‘ar-rijāl-u <i>the-men.MASC.PL-NOM.DEF</i> | katab-ū <i>wrote-MASC.PL</i> |
| ‘the men wrote’ | |
| ‘al-banāt-u <i>the-girls.FEM.PL-NOM.DEF</i> | ḍahik-na <i>laughed-FEM.PL</i> |
| ‘the girls laughed’ | |

However, if the verb precedes its nominal subject, it agrees in gender but not in number:

| | |
|---------------------------------------|--|
| katab-a <i>wrote-MASC.SING</i> | r-rijāl-u <i>the-men.MASC.PL-NOM.DEF</i> |
| ‘the men wrote’ | |
| ḍahik-at-i <i>laughed-FEM.SING</i> | l-banāt-u <i>the-girls.FEM.PL-NOM.DEF</i> |
| ‘the girls laughed’ | |

A special case of agreement occurs with the cardinal numbers from 3 to 10, which take the opposite gender of the counted noun’s singular, itself added in the genitive plural:

| | | |
|---------------------------------------|---|--|
| khams-u <i>five.MASC-</i> NOM | sanaw-āt-i-n <i>years-PL-GEN-</i> INDEF | [san-at-u-n] <i>[year-FEM.SING-NOM-</i> INDEF] |
| ‘five years’ | | |
| khams-at-u <i>five-FEM-</i> NOM | ‘ayyām-i-n <i>days.PL-GEN-INDEF</i> | [yawm-u-n] <i>[day.MASC.SING-</i> NOM-INDEF] |
| ‘five days’ | | |

Equational Sentences Positive equational sentences in the present have no copula:

| | |
|--|---|
| ‘al-bayt-u <i>the-house.MASC.SING-</i> NOM.DEF | kabīr-u-n <i>big.MASC.SING-NOM-</i> INDEF |
| ‘the house is big’ | |
| ‘anti <i>you.FEM.SING.NOM</i> | ṭabīb-at-u-n <i>physician-FEM.SING-NOM-INDEF</i> |
| ‘you (FEM) are a physician’ | |

For the negated present, the special verb *laysa* ‘to be not’ is used; in all other cases, appropriate forms of the verb *kān-a* ‘to be’ appear. Both verbs exhibit the peculiarity that their nominal complement is in the accusative:

| | | |
|------------------------------------|---------------------------------|-----------------------------------|
| ‘al-bayt-u <i>the-house-NOM</i> | laysa/kāna <i>is-not/was</i> | kabīr-a-n <i>big-ACC-INDEF</i> |
| ‘the house is not/was big’ | | |

Subordination Temporal, final, causative, and other clauses are usually introduced by subordinating conjunctions such as *lammā* ‘when’, *li-* ‘in order to’, and *li-‘anna* ‘because’.

Constructions with the conjunction *wa-* ‘and’ are frequently used to express simultaneousness of actions or events (in Arabic, called *ḥāl*, ‘circumstance’ sentence):

| | |
|--|--------------------------------------|
| dakhal-a <i>entered.PAST-3.SING</i> | l-ghurfata <i>the-room</i> |
| wa-huwa <i>and-he</i> | yaḍḥak-u <i>laughs.PRES-INDIC</i> |
| ‘he entered the room laughing’ | |

Relative Clauses In Arabic, relative clauses are complete sentences that are normally linked to their head by a personal pronoun referring to it. A relative pronoun, which agrees in number and gender, is used only if the head is definite:

| | | | |
|-----------------------------------|---------------------------|--------------------|----------------------------|
| ‘al-bint-u <i>the-girl-NOM</i> | llatī <i>REL.PRON.</i> | hiya <i>she</i> | faqīr-at-un <i>poor</i> |
| ‘the girl who is poor’ | | | |

'al-bint-u llatī ra'ay-tu-hā 'amsi
the-girl- NOM REL.PRON. SING.FEM saw-1.SING- her.ACC yesterday
 'the girl whom I saw yesterday'

'al-bint-u llatī māt-at
the-girl-NOM REL.PRON.SING.FEM died-3.FEM
 'umm-u-hā
mother-NOM-her.GEN
 'the girl whose mother has died'
 bint-u-n māt-at 'umm-u-hā
girl-NOM-INDEF died-3.FEM mother-NOM-her.GEN
 'a girl whose mother has died'

Syntax of the Dialects In principle, most dialects have preserved the combined time aspect system, although there are tendencies toward a stricter tense system (perfect for past, imperfect for nonpast). Frequently found as a discourse mechanism, however, is the narrative imperfect, in which a single past-time reference gives the frame for a following series of imperfective verb forms describing past actions or events. The perfect aspect expressed by the participle has become a well-established category in many, particularly eastern, dialects.

A great variety of auxiliary verbs (also called aspectualizers) exists for emphasizing punctual, durative, ingressive, progressive, and other aspects (see, for Cairo, Eisele, 1999).

Regarding word order, recent studies (Dahlgren, 1998; Brustad, 2000) have shown that the alleged preference of the dialects for SV order is true only for some urban dialects. On the whole, the same principles of thematization and foreground/background distinction obtain in the spoken vernaculars: "VSO represents the dominant typology in event narration, while SVO functions as topic-prominent typology that is used to describe and converse" (Brustad, 2000: 361). Particularly in dialogs, OVS is very frequent in topic-prominent structures, in which case a pronominal suffix has to mark the original place of the object, for example, in the Cairene dialect:

ukht-ak shuf-t-aha mbārīh
sister-your saw-1.SING-her yesterday
 'I saw your sister yesterday'

Agreement of nonhuman plural with feminine singular is possible, but in nearly all dialects 'logical' agreement is widely found. In the dialect of Damascus, both of the following phrases are equally acceptable: *byūt zghīre* ~ *byūt zghār* 'small houses'. Which of the two is used depends on semantic, idiomatic, and stylistic features insufficiently investigated in detail. In general, the word order has no influence on agreement; that is, a verb usually agrees with its nominal

subject in number whether the noun precedes or follows the verb.

Many dialects have developed so-called 'genitive exponents,' particles that are used under certain conditions for an analytic linking of two nouns or a noun and a pronoun suffix instead of a direct annexation. For example, in Arab Gulf dialect:

mēz māl ṭa'ām
table GEN PRT meal
 'dining table'

and in Cairo dialect:

il-baṣbūr bitā'i
the-passport GEN PRT-my
 'my passport'

Etymologically most of these particles can be traced back to a word meaning 'property' or 'right'. The choice whether an analytic or a synthetic construction is preferred depends on stylistic, semantic, and syntactical principles (Eksell Harning, 1980).

Bibliography

- Badawi E, Carter M G & Gully A (2004). *Modern written Arabic: a comprehensive grammar*. New York: Routledge.
 Brustad K E (2000). *The syntax of spoken Arabic*. Washington, DC: Georgetown University Press.
 Cantarino V (1974–1975). *Syntax of modern Arabic prose* (3 vols). Bloomington: Indiana University Press.
 Cuvalay-Haak M (1997). *The verb in literary and colloquial Arabic*. New York: Mouton de Gruyter.
 Dahlgren S-O (1998). *Word order in Arabic*. Gothenburg: Acta Universitatis.
 Eisele J C (1999). *Arabic verbs in time: tense and aspect in Cairene Arabic*. Wiesbaden: Harrassowitz.
 Eksell Harning K (1980). *The analytic genitive in modern Arabic dialects*. Gothenburg: Acta Universitatis.
 Faber A (1997). 'Genetic subgroupings of the Semitic languages.' In Hetzron R (ed.) *The Semitic languages*. New York: Routledge. 3–15.
 Ferguson C A (1959). 'The Arabic koine.' *Language* 25, 616–630.
 Fischer W (2001). *A grammar of classical Arabic* (3rd edn.). New Haven: Yale University Press.
 Fischer W & Jastrow O (1980). *Handbuch der arabischen Dialekte*. Wiesbaden: Harrassowitz.
 Fleisch H (1961–1979). *Traité de philologie arabe* (2 vols). Beirut: Dar el-Machreq.
 Holes C (1995). *Modern Arabic: structures, functions, and varieties*. New York: Longman.
 Miller A (1986). 'The origin of the modern Arabic sedentary dialects: an evaluation of several theories.' *Al-'Arabiyya* 19, 47–74.
 Murtonen A (1964). *Broken plurals: origin and development of the system*. Leiden: Brill.
 Peled Y (1992). *Conditional structures in classical Arabic*. Wiesbaden: Harrassowitz.

- Rabin C (1951). *Ancient west Arabian*. London: Taylor's Foreign Press.
- Suleiman Y (1994). *Arabic sociolinguistics: issues and perspectives*. Richmond, VA: Curzon.
- Veersteegh K (1997). *The Arabic language*. Edinburgh, UK: Edinburgh University Press.
- Watson J C E (2002). *The phonology and morphology of Arabic*. Oxford: University Press.
- Wehr H (1979). *A dictionary of modern written Arabic. Arabic-English* (4th edn.), Cowan J M (eds.). Wiesbaden: Harrassowitz.

Wright W W (1974). *A grammar of the Arabic language*. Beirut: Librairie du Liban.

Relevant Website

<http://semarch.uni-hd.de.SemiticSoundArchive>. University of Heidelberg.

Arabic as an Introflecting Language

J C E Watson, University of Oslo, Oslo, Norway

© 2006 Elsevier Ltd. All rights reserved.

In some languages, words are constructed or partially constructed not through the concatenation of linearly separable morphemes (e.g., English un-accept-able), but by the interdigitation of morphological forms which individually do not constitute self-standing phonological wholes. This type of morphology is variably termed in the literature introflectional, nonconcatenative (McCarthy, 1981), or transfixing (Bauer, 2003). It is a pervasive feature of the Semitic languages, and is particularly highly developed in Arabic. A simple example of introflection in Arabic is provided by *katab* 'wrote,' consisting of the root *k-t-b* {write}, the template CVCVC {PERF} and the vocalic melody a-a {ACT}.

Although introflection is a central feature of Arabic, most inflectional and some derivational categories are expressed through affixation; many derivational categories, which are expressed principally by introflection, take complementary prefixes or, less commonly, suffixes. This entry focuses on the morphology of Modern Standard Arabic, the formal written-based variety of the language, although many of the features outlined here are also found in the hundreds of Arabic dialects identifiable across the Arab world. The entry deals first with introflecting morphology in Arabic, sometimes in combination with affixation, and goes on to consider how introflection interacts with inflecting morphology.

Root and Pattern

Basic noun and verb stems in Arabic comprise a consonantal root and a pattern. The pattern can be further divided into two elements – a prosodic template and a vocalic melody. Most consonantal roots are trilateral. The root prototypically expresses the content

meaning of the word, the pattern functional meaning. The association of the consonantal root and vocalic melody with the prosodic template is illustrated for the verb stem *katab* 'wrote.ACT' in Figure 1.

The consonantal root is always fully independent of the prosodic template; the vocalic melody, by contrast, shows independence for relatively few morphological categories; such examples include *katab* 'wrote.ACT' versus *kutib* 'wrote.PASS' in which the vocalic melody alone expresses voice. However, in the word *ṣīlaaj* 'healing; treatment,' which comprises the consonantal root *ṣ-l-j* {heal; treat}, the prosodic template CVCVVC, and the vocalic melody *i-a*, the combination of the latter two expresses the category of verbal noun, rather than either the prosodic template or the vocalic melody independently.

Verbal Morphology

As illustrated in Table 1, Modern Standard Arabic has one basic verb form (form I) and nine derived forms (forms II–X), each of which typically imposes a more specific sense on that of the basic form: forms II, III, and IV are derived from form I by extension of

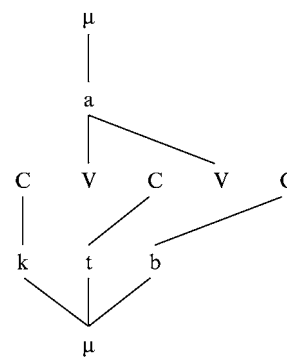


Figure 1 Association of consonantal root and vocalic melody.

Table 1 Verb forms I–X

| Typical meaning extension | Form | PERF ACT | Gloss | PERF PASS | IMPERF ACT | IMPERF PASS |
|---|------|----------|----------------------|------------|------------|-------------|
| | I | katab | write | kutib | yaktub | yuktab |
| causative | II | kattab | make s.o. write | kuttiib | yukattiib | yukattiab |
| attempt | III | kaatab | correspond with s.o. | kuuttiib | yukaattiib | yukaattab |
| causative | IV | ʔaktab | dictate | ʔuktiib | yuktiib | yuktab |
| reflexive of II | V | takattab | NA | tukuuttiib | yatakattab | yutakattab |
| reflexive of III | VI | takaatab | write to e.o. | tukuuttiib | yatakaatab | yutakaatab |
| medio-passive | VII | inkatab | subscribe | unkutiib | yankatiib | yunkatab |
| reflexive | VIII | iktatab | be recorded | uktutiib | yaktatiib | yuktatab |
| be/come a color/defect (e.g., red/lame) | IX | iktabb | NA | uktibb | yaktibb | yuktabb |
| reflexive of IV | X | istaktab | ask s.o. to write | ustutiib | yastaktiib | yustaktab |

the stem; forms V and VI are derived by prefixation of *ta-* to forms II and III, respectively. Forms VII, IX, and X involve various types of prefixation, and form VIII is derived from form I by infixation of *t* after the left-most root consonant. No consonantal root in Modern Standard Arabic has all ten verb forms, and a few verbs have one or more derived forms but lack the basic form. The prosodic template expresses the verbal form, the vocalic melody voice and aspect. The imperfect is distinguished from the perfect by imperfect person prefixes, and, in the case of form I only, by a different prosodic template. The root *k-t-b* {write} is used to illustrate verb forms in Table 1. The prototypical meaning correlates of the derived forms are listed in column two, and the specific meanings associated with the root *k-t-b*, where attested for the form in question, in column four.

The vocalic melody *a-a* indicates perfect aspect active voice, *u-i* perfect aspect passive voice, and *u-a* imperfect aspect passive voice. Excepting forms I, V and VI, the vocalic melody is *(u)-(a)-i* for the imperfect aspect active voice, and *(u)-(a)-a* for the imperfect passive. The same vocalic melodies express voice in the verbal participles, which are distinguished from the verb forms by the complementary prefixation of *mu-* to the stem. Active and passive participles from verb forms II–X are illustrated in Table 2.

Nominal Morphology

In contrast to participles from forms II–X, participles from form I verbs are derived through prosodic change: lengthening of the left-most vowel for the active participle, and of the right-most vowel for the passive participle, which also takes the complementary prefix *ma-*. Thus, *katab* ‘wrote’ has the participles *kaatiib* ‘writing; writer’ and *ma-ktuub* ‘written; letter.’

Finite verb stems are marked prosodically by a final light syllable – CVC, as seen in Table 1. As shown in Table 3, verbal nouns of most derived verbs (all tokens of forms IV, VII, VIII, IX, X, some of III), and a

Table 2 Active and passive participles

| Form | PART ACT | PART PASS |
|------|--------------|-------------|
| II | mukattiib | mukattab |
| III | mukaattiib | mukaattab |
| IV | muktiib | muktab |
| V | mutakattiib | mutakattab |
| VI | mutakaattiib | mutakaattab |
| VII | munkatiib | munkatab |
| VIII | muktatiib | muktatab |
| IX | muktibb | muktabb |
| X | mustaktiib | mustaktab |

Table 3 Verbal noun patterns

| Form | Verbal noun |
|------|----------------------|
| I | kitaab-ah |
| II | taktiib |
| III | kitaab / mukaatab-ah |
| IV | ʔiktaab |
| V | takattub |
| VI | takaatub |
| VII | inkitaab |
| VIII | iktitaab |
| IX | iktaabb |
| X | istiktaab |

number of form I verbs, are derived from finite verbs by lengthening of the stem-final syllable to CVVC and the vocalic melody *i-a*, the inverse of the vocalic melody for the active participle. Exceptions are form II, which has a complementary prefix *ta-* and the vocalic melody *a-i*, one form III variant (*mu-kaatab-ah*), and forms V and VI, both distinguished from the finite verb by umlaut of the stem-final vowel to *-u-*.

Singular Nouns and Adjectives

In contrast to verbs, singular nouns and adjectives take a vast array of different prosodic templates and vocalic melodies. Some, such as CaCCaaC, typically used for nouns of profession (e.g., *jazzaar*

‘butcher’), and the typically adjectival CaCuuC (e.g., *ḥasuud* ‘envious’), and CaCiiC (e.g., *kabiir* ‘big; old’), have a restricted range of meanings. Other patterns, such as CaCC, have a large range of meanings, covering human (*jadd* ‘grandfather’), non-human (*kalb* ‘dog’), concrete (*baḥr* ‘sea’), abstract (*ʿaql* ‘intelligence’), and adjectives (*ḥayy* ‘alive’).

Broken Plurals and Diminutives

Plurals are formed in Arabic in one of two ways: either through ‘sound’ plural suffixes or through the rich set of ‘broken’ plurals, wherein the plural is derived by mapping a portion of the singular to a plural prosodic template. McCarthy and Prince (1990a,b, 1998) have successfully analyzed broken plural derivation in moraic terms. The majority of singulars comprising three or more moras take predictable broken plural patterns. To derive the plural from such nonminimal singulars, the first two moras of the singular are mapped to an iambic template. *makaatib* ‘offices’ is derived from *maktab* ‘office,’ for example, as follows: the first two moras of the singular (*mak*) are mapped to an iambic template ($\mu\mu$) to give $m\mu k\mu\mu$. The vocalic melody *-a-* associates to the moraic slots to give *makaa*. The remainder of the singular (*-tab*) is suffixed to the iamb, and where this contains a vocalic slot, as here, *-i-* of the plural vocalic melody overrides the vowel of the remainder, to give *makaatib*. In the case of words comprising two moras and a number of non-minimal words, the plural cannot be predicted as easily from the singular form. Examples include *bayt* ‘house’ pl. *buyuut*, *bint* ‘girl’ pl. *banaat*, *kitaab* ‘book’ pl. *kutub*, *walad* ‘boy’ pl. *ʔawlaad*.

Whereas broken plural derivation is predictable in a proportion of cases, the diminutive is totally predictable and can, at least as far as Standard Arabic is concerned, be derived from almost any singular noun or adjective: the first two moras of the unmarked singular are mapped to an iambic template, as for the broken plural. From *walad* ‘boy,’ *wala* maps to $w\mu l\mu\mu$. The vocalic melody *u-ai* associates to the moraic slots to give *wulai*; the remainder of the singular (*-d*) is added, to derive *wulaid* ‘little boy’.

Elatives (Comparatives, Superlatives)

Elatives are derived predictably from most basic adjectives. The elative pattern is *ʔaCCaC* for trilateral roots. The vocalic melody (*-a-*) is dependent on the pattern. Examples include: *ʔakbar* ‘bigger; older’ (*kabiir* ‘big; old’); *ʔaʃʔab* ‘more difficult’ (*ʃaʔb* ‘difficult’); *ʔajban* ‘more cowardly’ (*jabaan* ‘cowardly’); *ʔaḥsan* ‘better’ (*ḥasan* ‘good’).

Inflectional Morphology

While stems are partially or wholly the product of introflection, grammatically complete words involve further affixation. Affixational elements include:

- Verbal pronominal prefixes and suffixes
- Object suffixes
- Possessive suffixes
- *-at-* feminine suffix
- Sound plurals
- Dual
- Case (nominative *-u*, accusative *-a*, genitive *-i*)
- *-n* suffix (indefinite/non-construct marker)
- Mood endings (indicative *-u*, subjunctive *-a*, jussive *-o*)

Pronominal prefixes and most suffixes, the feminine suffix, sound plurals and the dual comprise consonants and vowels, whereas all three case markers and indicative and subjunctive mood markers for the imperfect aspect are simple vowel endings. As seen in Table 4, pronominal subject markers are suffixal in the perfect aspect; in the imperfect aspect, pronominal markers are suffixal and/or prefixal. The jussive mood is given in Table 4 in the imperfect column. The indicative is expressed by suffixation of *-u* to forms ending in a root consonant (here *-b*) and suffixation of *-na* to forms ending in a vocalic suffix. The subjunctive is expressed by suffixation of *-a* to forms ending in a root consonant.

Sound Plural and Dual

Arabic has two nominal ‘sound’ plural suffixes: masculine and feminine. The sound feminine plural *-aat* takes the endings *-u* for nominative and *-i* for accusative or genitive case, and, further, *-n* to express indefiniteness or non-construct, as in:

mudarris-aat-u-n
teacher-FEM.PL.-NOM-INDEF
‘teachers FEM.PL.’

Table 4 Verbal inflections

| PERS/NUM/GEN | PERF | IMPERF.JUSSIVE |
|--------------|-------------|----------------|
| 1 s. | katab-tu | ʔ-aktub |
| 1 pl. | katab-naa | n-aktub |
| 2 s.m. | katab-ta | t-aktub |
| 2 s.f. | katab-ti | t-aktub-ii |
| 2 pl.m. | katab-tum | t-aktub-uu |
| 2 pl.f. | katab-tunna | t-aktub-na |
| 3 s.m. | katab-a | y-aktub |
| 3 s.f. | katab-at | t-aktub |
| 3 pl.m. | katab-uu | y-aktub-uu |
| 3 pl.f. | katab-na | y-aktub-na |

The sound masculine plural has two main forms: nominative *-uuna* and accusative/genitive *-iina*. The dual morpheme, suffixed to masculine or feminine nouns or adjectives, also has two main forms – *-aani* for the nominative and *-aini* for the accusative/genitive case.

Bibliography

- Bauer L (2003). *Introducing linguistic morphology*. Edinburgh University Press.
- Beeton A F L (1970). *The Arabic language today*. London: Hutchinson.
- Eid M & McCarthy J J (eds.) (1990). *Perspectives on Arabic linguistics II*. Amsterdam: John Benjamins.
- Fischer W (1997). 'Classical Arabic.' In Hetzron R (ed.) *The Semitic languages*. London: Routledge. 187–219.
- Gafos A I (2002). 'An argument for a stem-based view of Arabic morphology: Doubled verbs revisited.' In Parkinson D B & Benmamoun E (eds.) 59–86.
- Greenberg J H (1991). 'The patterning of root morphemes in Semitic.' *Word* 6, 162–181.
- McCarthy J J (1981). 'A prosodic theory of nonconcatenative morphology.' *Linguistic Inquiry* 12(3), 373–418.
- McCarthy J J (1994). 'Morphology, nonconcatenative.' In Asher R E (ed.) *The encyclopedia of language and linguistics*. Oxford: Pergamon. 2598–2600.
- McCarthy J J & Prince A (1990a). 'Prosodic morphology and templatic morphology.' In Eid M & McCarthy J J (eds.) 1–54.
- McCarthy J J & Prince A (1990b). 'Foot and word in prosodic morphology: The Arabic broken plural.' *Natural Language and Linguistic Theory* 8, 209–283.
- McCarthy J J & Prince A (1998). 'Prosodic morphology.' In Spencer A & Zwicky A M (eds.) *The handbook of morphology*. Oxford: Blackwell. 283–305.
- Parkinson D B & Benmamoun E (eds.) (2002). *Perspectives on Arabic linguistics XIII–XIV*. Amsterdam: John Benjamins.
- Ratcliffe R R (1990). 'Arabic broken plurals: Arguments for a two-fold classification of morphology.' In Eid M & McCarthy (eds.), 94–119.
- Ratcliffe R R (1998). *The 'broken' plural problem in Arabic and comparative Semitic: Allomorphy and analogy in non-concatenative morphology*. Amsterdam: John Benjamins.
- Ratcliffe R R (2002). 'The broken plural system of Moroccan Arabic: Diachronic and cognitive perspectives.' In Parkinson D B & Benmamoun E (eds.) 87–109.
- Versteegh K (1997). *The Arabic language*. Edinburgh University Press.
- Watson J C E (2002). *The phonology and morphology of Arabic*. Oxford University Press.

Arabic Languages, Variation in

E Al-Wer, University of Essex, Wivenhoe Park, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Arabic is a native language to some 200 million people, distributed over 22 different countries collectively known as the 'Arab World.' The Arab World stretches from the Indian Ocean in the east to the Atlantic Ocean in the west and includes most of the countries of the Middle East, the whole of North Africa, and Sudan (as well as Somalia and Mauritania). The Asian part of the Arab World is commonly referred to as al Mashreq 'the East,' and the North African part (particularly from and including Libya westwards) as al Maghreb 'the West.' Egypt represents the geographical link between the East and the West, and the Egyptian dialects may be thought of as a bridge between the Maghreb and the Mashreq dialects. In terms of demographic distribution, approximately 66% of the total population live in the African part. The largest concentration of Arabic speakers is in Egypt (67 million).

A Historical Sketch

The ancient home of the Arabs is the Arab Peninsula, and the Arabic language is traced to the second millennium B.C. in the northern part of the peninsula. To varying extents, everywhere else, Arabic is a relative newcomer. From the peninsula, and starting in the second half of the seventh century A.D., the language was disseminated first through direct military conquest, and later it affirmed its position through intellectual influence. In the course of its spread northwards to the eastern Mediterranean, Mesopotamia, and Egypt it ousted Greek, Persian, Aramaic, and Coptic. In the Maghreb, Arabic obscured Berber, and although it never managed to obliterate Berber, which continues to be spoken by no less than 40 million people, it altered the linguistic shape of the region.

Arabic prospered in a climate of dominant Arab civilization and declined alongside the diminution in power and influence of the Arabs. The rise to power of the Ottoman Turkish Empire in the 16th century resulted in the replacement of Arabic by Turkish as the language of state administration, although Turkish never managed to replace colloquial Arabic as the

everyday language of communication in the Arabic speaking provinces. The Ottomans lost the Maghreb in the mid-19th century (to Italy in Libya, and to France in Morocco, Algeria, and Tunisia), and Egypt to Napoleon for a short period of time and then to Britain. The outcome of the First World War brought an end to Ottoman rule in the Mashreq. Most of the Arabic speaking provinces were then divided into separate political entities and were placed under the tutelage of Britain and France. The linguistic significance of these developments was mainly that French and English became important features on the linguistic scene. English, however, did not influence the linguistic identity of the regions that came under British rule; it became at best the most widely spoken foreign language. French, on the other hand, had a far-reaching influence that continues to be visible, especially in the Maghreb, to this day. Much of the colonized or mandated territories became independent by the early 1960s, and Arabic has since then been declared the official language in Arab World.

Varieties of Arabic

To provide a concise outline of variation in Arabic, I will deal with two issues: the linguistic resources available to speakers of Arabic, and the sociolinguistic determinants of variation in Arab communities.

Standard Arabic

Throughout the Arab World, Standard Arabic (a modernized version of Classical Arabic), in an almost invariant form, is designated as the official language, the medium of instruction in education, and the language of the mass media, although in actual practice a mixture of Standard and colloquial varieties is used in education and in the media. The language was standardized twelve centuries ago, and the Standard variety has not been a spoken language for longer than that (see Holes, 1995a). It is not ordinarily used for everyday spoken purposes by any sector of the population. A functional knowledge of it is attainable through formal learning only, i.e., it is not acquired naturally. It stands in a diglossic relation to the spoken dialects (e.g., Spoken Egyptian Arabic), very much along the lines explained by Ferguson (1959).

The fact that this variety is not associated with a particular social group in contemporary Arab communities, and is not spoken natively, has sociolinguistic ramifications. There is no doubt that the Standard variety is accorded the highest status by Arabs, but its esteem and the degree to which it is involved in the course of linguistic change are unrelated. Research shows that linguistic variation and change in Arabic

is determined by interplay between local dialects and emerging local or regional standards, independently of Standard Arabic (see Al-Wer, 1997). Educated speakers of Arabic do resort to the use of Standard lexemes and constructions in formal situations. This is largely due to the established appropriateness of the Standard in such domains, and to the fact that learned and specialized lexical items are only available in a Standard form. Outside these situations, educated speakers use the colloquial varieties, and research shows that where linguistic change is in progress away from Standard features, the educated generally lead other groups, in the same way that they lead when the change happens to be in the direction of a Standard feature (for instance, see the results in Jabeur, 1987 and Al-Wer, 1991, and the discussion in Holes, 1995b).

The Dialects

Arabic dialects are the linguistic systems that speakers of Arabic speak natively. They vary considerably from region to region, with varying degrees of mutual intelligibility (and some are mutually unintelligible). Many aspects of the variability attested in the modern dialects can be found in the ancient Arabic dialects in the peninsula (for a detailed description of the ancient dialects, see El-Gindi, 1983). By the same token, many of the features that characterize various modern dialects, or distinguish between them, can be traced to the original settler dialects. In terms of typological classification, Arabic dialectologists distinguish between two basic norms: Bedouin and Sedentary. This classification is based on a bundle of phonological, morphological, and syntactic features that distinguish between the two norms. In the modern, especially urban dialects, it is not really possible to maintain this classification, partly because the modern dialects are typically an amalgam of features from both norms. Geographically, modern Arabic dialects are classified into five groups: Arabian Peninsula (four subgroups); Mesopotamian; Syro-Lebanese (or Levantine, three subgroups); Egyptian (four subgroups); and Maghreb (two subgroups) (for details, see Versteegh, 1997).

Common Dimensions of Variation in Arab Communities

There is a general shortage of studies on variation in Arabic, especially on Arabic in its social setting and in large and heterogeneous urban environments; but this situation is changing. A number of important empirical research studies, utilizing modern methodological and analytical techniques, are in preparation.

On the basis of the studies available, it seems that the factors outlined below play important roles in the dynamics of variation and the course of linguistic change.

All variation studies on Arabic mention education as an important social variable, and indeed the findings show that linguistic usage correlates with the level of education of speakers. However, the exact denotation of education as a variable is poorly understood. It is noticeable, for instance, that while level of education of the speaker is used as a sampling tool, it is not integrated in the explanatory model in a consistent way. It is likely that this variable actually symbolizes different aspects of the speakers' characteristics in different communities. It is also likely to be a proxy variable, acting on behalf of such things as contact and exposure to outside communities, especially since in many communities institutions of education are not available locally, and generally the longer individuals spend in formal education the more frequent their contacts become with speakers of other dialects (see Al-Wer, 2002a). In some cases, the type of education, private or public, was found important (as in Haeri's 1997 study in Cairo).

Social class is not usually used in Arabic studies. A notable exception in this domain is Haeri's (1997) study in Cairo, which analyzes this variable and finds it significant. A forthcoming study on Damascus also uses social class as a sampling and analytical tool. There are two types of urban Arabic communities, which seem to show different correlational patterns. In the well-established urban centers, such as Cairo and Damascus, the original regional, ethnic, or sectarian linguistic distinctions among the population are blurred and do not play a role in sociolinguistic correlations. On the other hand, in the new cities, such as Amman (the capital city of Jordan) and most of the cities in the Gulf region, stratification along ethnic, regional, and sectarian backgrounds are the more relevant criteria for sociolinguistic studies. There are signs that as these cities become established and their new dialects become focused, alternative ways of stratification become necessary. For instance, in Amman, the original distinctions of Jordanian versus Palestinian dialects and urban versus rural Palestinian (which are based on the regional origins of the city's population), while continuing to be important for an understanding of patterns of linguistic variation among certain groups, are much less important in the speech of the third generation inhabitants of the city. Other, more locally defined criteria, such as socioeconomic class, are becoming significant (for more details, see Al-Wer, 2002b).

Gender has been found to be an important parameter of variation in Arabic. Consistent linguistic

differences between male and female speakers are reported in the earlier studies (e.g., Abdel-Jawad, 1981, and Bakir, 1986), as well as in later works (e.g., Jabeur, 1987; Haeri, 1997; Gibson, 1998). Gender is also reported to be significant in studies focusing on code switching and code mixing (e.g., Lahlou, 1992, and Sadiqi, 2002). The interpretation of gender-differentiated patterns in Arabic experienced a complete transformation, although the patterns themselves are consistent and are in keeping with the patterns found in other languages, such as English. In the earlier studies, Arabic was thought to contravene the then generally reported tendency for female speakers to use standard features more often than men, since in Arabic studies men were found to use Standard Arabic features more than women. However, the features that Arab men were found to use more often than Arab women were at the same time characteristic of the localized and in many cases overtly stigmatized varieties, but simply happened to be identical to Standard Arabic features. Since the approach to understanding variation in Arabic has shifted from one based on the assumption that approximation to Standard Arabic features is the governing factor to one recognizing that the target features are characteristic of the *de facto* spoken local standards (which derive their status from the social groups whose speech they represent), the interpretation of gender patterns has also shifted (see Ibrahim, 1986; Haeri, 1987; Al-Wer, 1997).

Within this revised framework, the findings with respect to male-female differences in Arabic communities studied so far suggest that where linguistic change is in progress, allowing for other factors, the female speakers are ahead of the male speakers in the use of newer forms. However, it must be emphasized that the data available from Arabic do not permit us to make generalizations on the basis of gender (to the extent such generalizations can be made for any language). Although there is now a respectable number of sociological studies, mainly in the feminist literature, providing thorough analyses of gender as a social construct in Arab societies, these models have not yet been integrated in studies on linguistic variation in Arabic.

The current generation of students of Arabic linguistics increasingly pays attention to the study of dialect contact. This comes in recognition of the linguistic repercussions of the massive population movements, rapid urbanization, and modernization all over the Arab World. In the established cities, the newcomers largely accommodate to the city dialect (see for instance the results in Jabeur, 1987; Gibson, 1998; Jassem, 1993). In the new cities, various processes of

leveling take place and new linguistic forms emerge. There are also signs that regional koineization, transcending political borders, is taking place.

Bibliography

- Abdel-Jawad H (1981). *Lexical and phonological variation in spoken Arabic in Amman*. Ph.D. diss., University of Pennsylvania.
- Al-Wer E (1997). 'Arabic between reality and ideology.' *International Journal of Applied Linguistics* 7, 251–260.
- Al-Wer E (2002a). 'Education as a speaker variable.' In Rouchdy A (ed.). 41–53.
- Al-Wer E (2002b). 'Jordanian and Palestinian dialects in contact: vowel raising in Amman.' In Jones M & Esch E (eds.) *Language change: the interplay of internal, external and extra-linguistic factors*. Berlin: Mouton de Gruyter. 63–79.
- Bakir M (1986). 'Sex differences in the approximation to Standard Arabic: a case study.' *Anthropological Linguistics* 28(1), 3–9.
- El-Gindi A (1983). *Allahajaat l'arabiyya fit turaath* [heritage of Arabic dialects]. Cairo: Addaar l'arabiyya lilkitaab.
- Ferguson C (1959). 'Diglossia.' *Word* 15, 325–340.
- Gibson M (1998). *Dialect contact in Tunisian Arabic: sociolinguistic and structural aspects*. Ph.D. diss., University of Reading, UK.
- Haeri N (1987). 'Male/female differences in speech: an alternative interpretation.' In Denning *et al.* (eds.) *Proceedings of the Fifth Annual Conference on New Ways of Analyzing Variation*. Stanford, CA: Stanford University.
- Haeri N (1997). *The sociolinguistic market of Cairo*. London: Paul Kegan International.
- Holes C (1995a). *Modern Arabic*. New York: Longman.
- Holes C (1995b). 'Community, dialect and urbanization in the Arabic speaking Middle East.' *BSOAS* 58, 270–287.
- Ibrahim M (1986). 'Standard and prestige language: a problem in Arabic sociolinguistics.' *Anthropological Linguistics* 28(1), 115–126.
- Jabeur M (1987). *A sociolinguistic study in Tunisia: Rades*. Ph.D. diss. University of Reading, UK.
- Jassem Z (1993). *Impact of Arab-Israeli wars on language and social change in the Arab World: the case of Syrian Arabic*. Kuala Lumpur: Pustaka Antara.
- Lahlou M (1992). *A morpho-syntactic study of code-switching between Moroccan Arabic and French*. Ph.D. diss., Austin, University of Texas.
- Rouchdy A (ed.) (2002) *Language contact and language conflict in Arabic*. USA and Canada: Routledge Curzon.
- Sadiqi F (2002). 'The language of introduction in the city of Fes: the gender-identity interaction.' In Rouchdy A (ed.). 116–132.
- Versteegh K (1997). *The Arabic language*. Edinburgh: Edinburgh University Press.

Aramaic and Syriac

J Watt, Cardiff University, Cardiff, UK

© 2006 Elsevier Ltd. All rights reserved.

Origin and Expansion

Aramaic is the native name of a language that first manifests itself in inscriptions in Syria early in the 1st millennium B.C. but that in subsequent centuries, during the period of the Assyrian and Persian empires, was widespread throughout the Near East and is found as far afield as Egypt, Cilicia, and Iran. Following the conquests of Alexander the Great, and during the subsequent eras of Macedonian and Roman influence, it co-existed with Greek as a principal medium of written communication over this wide area. The conquest of the region by the Arabs in the 7th century A.D. eventually brought its dominant position to an end, but it remained significant for many years thereafter as a spoken and, especially, a literary language. Greek writers designated it

Syriac, a term derived from *Assyrian*, and the Greek name was frequently preferred over the native one by Aramaic-speaking Jews and Christians, among whom *Aramaean* became a designation for their pagan neighbors. Over time, Aramaic developed a number of clearly distinct literary dialects, each evolving out of a local form of the language, and these were extensively employed by Jewish, Christian, and other religious communities. In contemporary usage, Syriac usually refers to the principal literary dialect employed by Christians, whereas Aramaic is retained as a generic term for the whole group. Spoken forms of the language have survived to this day among the religious communities that have preserved it in their liturgies and in a few places as an everyday language.

Aramaic belongs to the Semitic group of languages and, more particularly, to the northwest branch, which, according to prevalent opinion, contained in the 1st millennium B.C. two distinct strands, Canaanite (which includes Hebrew and Phoenician) and Aramaic. Despite its extensive use

in the Assyrian and Persian empires, it has left few literary or epigraphic remains from these periods, although those that have survived are of considerable importance in the study of the history of the region and fresh discoveries are steadily adding to the stock. In the Hellenistic and Roman periods, the material becomes more abundant, and, especially from the 4th century onward, a large body of Christian literature is preserved in Syriac and a substantial body of Jewish literature in Palestinian and Babylonian Aramaic dialects. Smaller extant corpora in dialects employed by the religious communities in question stem from the Mandaean, Samaritan, and Syro-Palestinian Christians who adhered to the Orthodox confession of the Byzantine emperors.

Phases and Dialects

Over the 3000 years of its recorded history, the language has naturally undergone many developments. Three broad phases are easily discernible. In the first, represented mainly in inscriptions and papyri, the form of the language is surprisingly uniform, and the differences between documents from diverse times and places are relatively minor. In the second, represented in the literature (beginning in the 4th century A.D. or earlier) of Jews, Christians, Samaritans, and Mandaeans, more marked dialectal differences are apparent, with two broad groupings. The eastern group, of Mesopotamian provenance, comprises Syriac, Jewish Babylonian Aramaic, and Mandaic; the western group comprises Jewish Palestinian Aramaic, Samaritan, and Syro-Palestinian Christian Aramaic. A third phase of modern eastern and western dialects can be discerned from approximately the 17th century. Further differentiation beyond these three is less clear cut, but in recent years a fivefold classification has gained considerable support, with the period prior to the emergence of the literature in the eastern and western literary dialects divided into three. The distinction between the earliest of these, Old Aramaic, and its successor is relatively unproblematic. Old Aramaic inscriptions belong to the period of the independent Aramaean states (10th–8th centuries B.C.) and exhibit a number of distinctive grammatical features, some of them similar to those known in Canaanite. Texts in Aramaic from the subsequent period come from a vastly greater area, but despite their wide geographical and chronological range they exhibit a high degree of homogeneity. Many of them are administrative in nature, and the language in which they are composed was evidently employed as a formal means of communication in much of the Assyrian, Babylonian,

and Achaemenid empires. Its adoption by the imperial chancelleries (a striking example of which is the presence of Aramaic ideograms in Pahlavi texts) is no doubt the reason for the high degree of standardization, and this phase is therefore commonly designated Official Aramaic. More problematic is the characterization of the Aramaic material originating between the end of the Achaemenid Empire and the beginnings of the extensive literature in the later Jewish dialects and classical (Christian) Syriac. In this period (roughly 200 B.C.–200 A.D.), several different dialects emerge in a number of localities. These include Palmyrene, Nabataean, Hatran, and Old Syriac (Edessan) inscriptions, and inscriptions and fragments of literary works from Palestine. Although all these dialects are quite similar to Official Aramaic and developed out of it, the influence of spoken local dialects or other languages (Arabic among the Nabataeans and Akkadian in Mesopotamia) led to a fragmentation and modification of the earlier fairly uniform Official Aramaic. In none of these areas, however, do we have evidence this early of the emergence of a vigorous or widespread new literary Aramaic. Although some scholars therefore consider this period as still belonging to the literary phase of Official Aramaic, others are sufficiently impressed by the differences to classify it as a new phase, Middle Aramaic, falling between Official Aramaic and the Later Aramaic of rabbinical Jewish and Christian Syriac literature. The expansion and consolidation of these religions was presumably responsible for the transformation of local dialects into significant and widespread means of literary expression.

Dialects and Religious Communities

The Aramaic inscriptions of Old, Official, and Middle Aramaic provide important information on deities worshipped in Syria and Mesopotamia in pre-Christian and early Christian times. Papyri from Egypt constitute the largest body of material in Official Aramaic, among which those of a Jewish military colony at Elephantine are of particular interest for the light they shed on the religious beliefs and practices of this group of Jews in the Achaemenid Empire. The language of the Aramaic sections of *Ezra* and *Daniel* also belongs to Official Aramaic and differs only slightly from that of the Elephantine papyri. Subsequent Jewish writings and inscriptions of Palestinian provenance belong to the Middle Aramaic phase and include fragments of a number of literary works preserved among the Dead Sea Scrolls. The problem of determining the form of

spoken Aramaic current in 1st century A.D. Palestine has attracted much attention on account of its relevance to New Testament studies. Although the literary and epigraphic material from the period is consistent with the use of Middle Aramaic, the extant material is still fairly sparse, and some scholars still hold (as did most of those of earlier generations, to whom Middle Aramaic was unknown) that the Palestinian dialects of later rabbinical literature are a valuable source for the reconstruction of the spoken language of the 1st century A.D.

The rabbinical literature in Jewish Palestinian Aramaic comprises various Targumim (paraphrastic Aramaic versions of sections of the Hebrew Bible), Midrashim (commentaries on the biblical books), and parts of the Palestinian Talmud. The latter two are written partly in Hebrew and partly in Aramaic. The Targum on the Pentateuch attributed to Onkelos and that on the *Prophets* attributed to Jonathan were used in Babylonia and, unlike the Palestinian Pentateuch Targum, do not therefore represent a purely Palestinian form of Aramaic. Jewish Babylonian Aramaic is represented in the Aramaic sections of the Babylonian Talmud and in the Responsa literature of the 8th–10th centuries A.D., the replies of the heads of the Babylonian academies to legal questions from scattered Jewish communities. The famous 13th century mystical work from Spain known as the *Zohar* is also written partly in Hebrew and partly in an artificial Aramaic. Samaritan Aramaic is represented principally by the Samaritan Targum to the Pentateuch and the theological treatise known as *Memar Marqah*, an important source for the knowledge of Samaritan religion. The language of the Mandaean texts is Eastern Aramaic, but linguistic as well as historical arguments have been advanced in favor of a Palestinian origin contemporary with the beginnings of Christianity. Because, however, these are not decisive, and the Mandaeans are only known in Iraq and further east, a Mesopotamian origin of the religion and the texts is still widely accepted.

The largest extant corpus of Aramaic literature is that in Syriac. Originally the local dialect of Edessa (modern Urfa), Syriac was adopted as a literary language by Christians throughout the Near East. Once adopted, it remained remarkably stable in most respects, although two slightly differing dialects (eastern and western, using different scripts and differing in the pronunciation of some vowels) emerged around the 5th century. These were associated respectively with the East Syrian Church (in Sasanid domains) and the Syrian Orthodox Church (in the Roman domains). Syriac-speaking Christians were active in the translation of Greek writings into Syriac,

not only the Bible and Greek patristic writers but also (from the 6th century) medicine (Galen) and logic (Aristotle). Their expertise in these secular subjects in the period of the 'Abbasid caliphate, and their ability to read both the relevant Greek texts and the earlier Syriac translations of them, stimulated Muslim interest in these subjects and led to the Syrians being in great demand as translators from Greek to Arabic, such translations being frequently done through a Syriac intermediary. Greek loanwords, grammatical forms modeled on Greek, and Greek syntax all greatly influenced Syriac, increasingly so from the 6th century. By contrast, the influence of Arabic on the literary language was slight. In the earlier period, the most striking literature in Syriac is the religious poetry of Saint Ephrem, which was much admired and imitated even beyond the Syriac language area. From the 10th century, Arabic replaced Syriac among Christians as the chief language of theology, philosophy, and medicine, but the 13th century saw a veritable West Syriac renaissance, embodied especially in the great polymath Bar Hebraeus, who wrote with equal facility in Syriac and Arabic. In contrast to the wide use of Syriac, Syro-Palestinian Christian Aramaic (alternatively designated Syro-Palestinian Syriac because it was written in the West Syriac script) was employed only in Palestine and Syria, and the extant texts (mostly biblical, liturgical, or hagiographical) are all translations from Greek.

Spoken Aramaic dialects have been in continuous use in a number of places right into modern times. Modern Western dialects of Aramaic are spoken, by Christians and Muslims, in three villages north of Damascus, namely Ma'lula, Bah'a, and Jubb 'Addin. Eastern dialects have been more extensively used by Christians in various localities. In the mountainous area of Southeast Turkey known as Tur 'Abdin, Turoyo ('the mountain language') is spoken by members of the Syrian Orthodox Church. Other Eastern Aramaic dialects have been spoken in modern times by the Jews of Kurdistan and Azerbaijan, most of whom have now emigrated to Israel, and a modern Mandaic dialect has survived in Iran. The greatest use of Aramaic in modern times, however, has been by East Syrian Christians, among whom a number of East Aramaic dialects have been employed. Modern literary Syriac (Swadaya) may be said to have begun with the printing of books in the local dialect by the American Presbyterian Mission at Urmia in Northwest Iran. Although the number of people currently using some form of Aramaic is small, their determination to keep it alive is a testimony to their pride in a language whose demonstrable lifespan extends to 3000 years.

Bibliography

- Brock S P (1989). 'Three thousand years of Aramaic literature.' *Aram* 1(1), 11–23.
- Brock S P (1996). *Syriac studies: a classified bibliography, 1960–1990*. Kaslik, Lebanon: Université Saint-Esprit de Kaslik.

- Fitzmyer J A (1979). 'The phases of the Aramaic language.' In Fitzmyer J A (ed.) *A wandering Aramean: collected Aramaic essays*. Missoula, MT: Scholars Press. 57–84.
- Fitzmyer J A & Kaufman S A (1992). *An Aramaic bibliography*. Baltimore and London: Johns Hopkins University Press.

Arawak Languages

A Y Aikhenvald, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

The Arawak language family contains the largest number of languages in Latin America. Geographically, it spans four countries of Central America – Belize, Honduras, Guatemala, Nicaragua – and eight of South America – Bolivia, Guyana, French Guiana, Surinam, Venezuela, Colombia, Peru, Brazil (and also formerly Argentina and Paraguay).

There are about 40 living Arawak languages. The first Native American peoples encountered by Columbus – in the Bahamas, Hispaniola, and Puerto Rico – were the Arawak-speaking Taino. Their language became extinct within a hundred years of the invasion. Spanish and many other European languages inherited a number of loans from Arawak languages. These include widely used words such as *hammock*, *tobacco*, *potato*, *guava*, and many other names for flora and fauna.

The creation of a mixed language of Arawak/Carib origin in the Lesser Antilles is one of the most interesting pieces of evidence on language history in pre-conquest times. Speakers of Iñeri, a dialect of the Arawak language now (misleadingly) called Island Carib, were conquered by Carib speakers. They developed a mixed Carib/Arawak pidgin that survived until the 17th century (Hoff, 1994). Speech of men and speech of women were distinguished in the following way. Women used morphemes and lexemes of Arawak origin, while men used lexical items of Carib origin and grammatical morphemes mostly of Arawak origin. The pidgin coexisted with Carib used by men and Iñeri used by women and children; it belonged to both parties and served as a bridge between them. This diglossia gradually died out with the spread of competence in Island Carib among both men and women. As a result, Island Carib, an Arawak language, underwent strong lexical and, possibly, grammatical influence from Carib.

The languages in areas settled by the European invaders soon became extinct. Those on the north coast of South America perished first, before 1700. When the search for gold and rubber extended up the Amazon and its tributary the Rio Negro, further languages succumbed, from the 18th century up until the present day. Sometimes the Indians retaliated, attacking settlements and missions; but the invaders always returned. Indian rebellions often provoked forced migrations which sometimes ended up creating a new dialect or even a language. For instance, in 1797 the British authorities removed the rebellious inhabitants of St. Vincent (an island in the Lesser Antilles) to Belize on the mainland. These were racially a mixture of black slaves and Indians, who spoke Island Carib. This resulted in the creation of a new dialect of Island Carib – known as Central American Island Carib, Kari'f, Black Carib, or Garifuna – which by the 20th century had developed into a separate language, now spoken in Central America (Taylor, 1977).

The overwhelming majority of Arawak languages are endangered. Even in the few communities with more than 1000 speakers, a national language (Portuguese or Spanish) or a local lingua franca (Língua Geral Amazônica, Quechua, or Tucano) is gaining ground among younger people. The few healthy Arawak languages are Guajiro in Venezuela and Colombia (estimates vary from 60 000 to 300 000 speakers) and the Campa languages (total estimate 40 000 to 50 000 speakers), one of the largest indigenous groups in Peru.

Most of the materials on Arawak languages collected during the second half of the 20th century are by missionary linguists. Their quality and quantity varies. Only three or four languages have full descriptions available.

The genetic unity of Arawak languages was first recognized by Father Gilij as early as 1783. The recognition of the family was based on a comparison of pronominal cross-referencing prefixes in Maipure, an extinct language from the Orinoco Valley, and in Moxo from Bolivia. Gilij named the family Maipure. Later, it was renamed Arawak by Daniel Brinton after

one of the most important languages of the family, Arawak (or Lokono), spoken in the Guianas. This name gained wide acceptance during the following decades. The majority of Native South American scholars use the name Arawak (Aruák) to refer to the group of unquestionably related languages easily recognizable by pronominal prefixes such as *nu-* or *ta-* ‘first person singular’, *(p)i-* ‘second person singular’, prefix *ka-* meaning ‘have’, and negator *ma-*. A number of scholars, mainly North Americans, prefer to use the term Arawak(-an) to refer to a much more doubtful higher-level grouping, and reserve the term Maipuran (or Maipurean) for the group of undoubtedly related languages that are claimed to be one branch of Arawakan (see Payne, 1991). Here I follow the South American practice and use the name *Arawak* for the family of definitely related languages.

The limits of the family were established by the early 20th century. Problems still exist concerning internal genetic relationships within the family and possible genetic relationships with other groups. Reconstruction, internal classification, and subgrouping of Arawak languages remain matters of debate; further detailed work is needed on both the descriptive and comparative fronts.

The putative studies of Arawakan by Ester Matteson, G. Kingsley Noble, and others are deeply flawed. Unfortunately, these have been adopted as the standard reference for the classification of Arawak languages, especially among some anthropologists, archaeologists, and geneticists, influencing ideas on a putative proto-home and migration routes for proto-Arawakan – see the criticism in Tovar and De Tovar (1984), Dixon and Aikhenvald (1999: 12–15), and Aikhenvald (1999a).

Little is known about a proto-home for the Arawak family. The linguistic argument in favor of an Arawak proto-home located between the Rio Negro and the Orinoco rivers – or on the Upper Amazon – is based on the fact that there is a higher concentration of structurally divergent languages found in this region. This area has also been suggested as one of the places where agriculture developed. This is highly suggestive and corroborated by a few mythical traditions of northern origin by Arawak-speaking peoples south of the Amazon. The origin myths of the Tariana, in northwest Amazonia, suggest that they could have come from the north coast of South America.

Arawak languages are complicated in many ways. Words can be differentiated by stress in some languages, such as Baure and Waurá (south of Amazonas), and Tariana, Achagua, and Warekena (north of Amazonas). At least two have tones – Terêna in the South, and Resígaro spoken in the far northeast of Peru.

Each Arawak language has a few prefixes and numerous suffixes. Prefixes are typically monosyllabic, while suffixes can consist of one or more syllables. Roots usually contain two syllables. Prefixes are rather uniform across the family, while suffixes are not. What is a free morpheme in one language can be a grammatical marker in another language; for instance, postpositions become causative markers, and nouns become classifiers. An Apurina noun *maka* means ‘clothing’ – this is where the word for hammock comes from. In Baniwa of Içana, *-maka* is a classifier for stretchable thin extended objects, e.g., *tsaia* ‘skirt’ or *dzawiya* ‘jaguar’s skin’, as in *apa-maka* (one-CLASSIFIER:CLOTHING) ‘one piece of clothing’.

Most grammatical categories in Arawak languages are verbal. Cases to mark subjects and objects are atypical. Tariana, spoken in northwest Brazil, has developed cases for core grammatical relations to match the pattern in nearby Tucanoan languages (Aikhenvald, 1999b).

Arawak languages spoken south of the Amazon (South Arawak) have a more complex predicate structure than those north of the Amazon (North Arawak). South Arawak languages such as Amuesha or Campa have up to thirty suffix positions. North Arawak languages such as Tariana or Palikur have not more than a dozen suffixes. Suffixes express meanings realized by independent words in familiar Indo-European languages, e.g., ‘be about to do something’, ‘want to do something’, ‘do late at night’, ‘do early in the morning’, ‘do all along the way’, ‘in vain’, ‘each other’.

Verbs are typically divided into transitive (e.g., ‘hit’), active intransitive (e.g., ‘jump’) and stative intransitive (e.g., ‘be cold’). All Arawak languages share pronominal affixes and personal pronouns. Pronominal suffixes refer to subjects of stative verbs and direct objects. Prefixes are used for subjects of transitive verbs and of intransitive active verbs, and for possessors. That is, most Arawak languages are of active-stative type. For instance, in Baniwa one says *nu-kapa* ‘I see’ and *nu-watsa* ‘I jump’, but *nu-kapa-ni* ‘I see him’ and *hape-ni* ‘he is cold’ (*nu-* refers to ‘I’ and *-ni* to ‘him’). And ‘my hand’ is *nu-kapi*.

Some languages have lost the pronominal suffixes (and with them the morphological basis for an active-stative system); these include Yawalapiti (Xingú area, Brazil) and Chamicuro (Peru) to the south of the Amazon, and Bare, Resígaro, Maipure, and Tariana to the north. The form of the first person pronoun is *ta-* in the Caribbean (Lokono, Guajiro, Añun, Taino) and *nu-* in other languages. This is the basis for classification of Arawak languages into *Nu-Arawak* and *Ta-Arawak*.

Table 1 Pronominal prefixes and suffixes in proto-Arawak

| Person | Prefixes | | Suffixes | |
|--------------|--------------------------|--------------|--------------------------------------|------------|
| | Singular | Plural | Singular | Plural |
| 1 | <i>nu-</i> or <i>ta-</i> | <i>wa-</i> | <i>-na</i> , <i>-te</i> | <i>-wa</i> |
| 2 | <i>(p)i-</i> | <i>(h)i-</i> | <i>-pi</i> | <i>-hi</i> |
| 3nf | <i>r i-</i> , <i>i-</i> | <i>na-</i> | <i>-r i</i> , <i>-i</i> | <i>-na</i> |
| 3f | <i>thu-</i> , <i>ru-</i> | <i>na-</i> | <i>-thu</i> , <i>-ru</i> , <i>-u</i> | <i>-na</i> |
| 'impersonal' | <i>pa-</i> | — | — | — |

Proto-Arawak must have had an unusual system of four persons: first, second, third, and impersonal. The forms of prefixes and suffixes reconstructed for proto-Arawak are given in Table 1.

Most Arawak languages distinguish two genders – masculine and feminine – in cross-referencing affixes, in personal pronouns, in demonstratives, and in nominalizations, e.g., Palikur *amepi-yo-* ‘thief (woman)’, *amepi-ye* ‘thief (man)’, Tariana *nu-phe-ri* ‘my elder brother’, *nu-phe-ru* ‘my elder sister’. No genders are distinguished in the plural. The markers go back to proto-Arawak third person singular suffixes and prefixes: feminine *(r)u*, masculine *(r)i*. Some languages also have complicated systems of classifiers – these characterize the noun in terms of its shape, size, and function (Aikhenvald, 1999a). For instance, Tariana and Baniwa of Içana have more than 40 classifiers which appear on numerals, adjectives, verbs, and in possessive constructions. Palikur has more than a dozen classifiers which have different semantics and form depending on whether they are used on numerals, verbs, or on adpositions (Aikhenvald and Green, 1998). Pronominal genders have been lost from some languages, e.g., Terêna, Amuesha, Chamicuro, Pareci, Waurá (south of the Amazon), and Bahwana (north of the Amazon).

All Arawak languages distinguish singular and plural. Plural is only obligatory with human nouns. Plural markers are **-na/-ni* ‘animate/human plural’, **-pe* ‘inanimate/animate non-human plural’. Dual number is atypical. In Resígaro, markers of dual were borrowed from the neighboring Bora-Witoto languages.

Throughout the Arawak language family, nouns divide into those which must have a possessor (inalienably possessed) and those which do not have to have a possessor (alienably possessed). Inalienably possessed nouns are body parts, kinship terms, and a few others, e.g., ‘house’ and ‘name’. Inalienably possessed nouns have an ‘unpossessed’ form marked with a reflex of the suffix **-<i* or **-hV*, e.g., Pareci *no-tiho* ‘my face’, *tiho-ti* ‘(someone’s) face’; Baniwa *nu-hwida* ‘my head’, *i-hwida-ti* (INDEFINITE-head-NON.POSSESSED) ‘someone’s head’. Alienably possessed nouns take

one of the suffixes **-ne/-ni*, **-te*, **-re*, **-i/-e* (Payne, 1991: 378), or **-na* when possessed, e.g., Baniwa *nu-<inu-ni* (1sg-dog-possessive) ‘my dog’.

The overwhelming majority of Arawak languages have a negative prefix *ma-* and its positive counterpart, prefix *ka-*, e.g., Piro *ka-yhi* (ATTRIBUTIVE-tooth) ‘having teeth’, *ma-yhi* (NEGATIVE-tooth) ‘toothless’; Bare *ka-witi-w* (ATTRIBUTIVE-eye-FEMININE) ‘a woman with good eyes’, *ma-witi-w* ‘a woman with bad eyes; a blind woman’.

The common Arawak lexicon (cf. Payne, 1991) consists mostly of nouns. There are quite a few body parts, fauna, flora, and artifacts. Only a few verbs can be reconstructed, e.g., **kau* ‘arrive’, **p^(da)* ‘sweep’, **po* ‘give’, **(i)ya* ‘cry’, **kama* ‘be sick, die’; **itha* ‘drink’. Most languages have just the numbers ‘one’ (proto-Arawak **pa-*; also meaning ‘someone, another’) and ‘two’ (proto-Arawak **(a)pi* and **yama*). A preliminary reconstruction is in Payne (1991). An up-to-date overview of the family is in Aikhenvald (1999a, 2001), and an overview of the proto-language is in Aikhenvald (2002).

Bibliography

- Aikhenvald A Y (1999a). ‘The Arawak language family.’ In Dixon R M W & Aikhenvald A Y (eds.). 65–105.
- Aikhenvald A Y (1999b). ‘Areal diffusion and language contact in the Içana-Vaupés basin, North West Amazonia.’ In Dixon R M W & Aikhenvald A Y (eds.). 385–415.
- Aikhenvald A Y (2001). ‘Areal diffusion, genetic inheritance and problems of subgrouping: a North Arawak case study.’ In Aikhenvald A Y & Dixon R M W (eds.). *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press. 167–194.
- Aikhenvald A Y (2002). *Language contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald A Y & Green D (1998). ‘Palikur and the typology of classifiers.’ *Anthropological Linguistics* 40, 429–480.
- Dixon R M W & Aikhenvald A Y (1999). ‘Introduction.’ In Dixon R M W & Aikhenvald A Y (eds.). 1–21.
- Dixon R M W & Aikhenvald A Y (eds.) (1999). *The Amazonian languages*. Cambridge: Cambridge University Press.
- Hoff B (1994). ‘Island Carib, an Arawakan language which incorporated a lexical register of Cariban origin, used to address men.’ In Bakker P & Mous M (eds.). *Mixed languages. 15 case studies in language intertwining*. Amsterdam: IFOTT. 161–168.
- Payne D L (1991). ‘A classification of Maipuran (Arawakan) languages based on shared lexical retentions.’ In Derbyshire D C & Pullum G K (eds.). *Handbook of Amazonian languages*, vol. 3. Berlin: Mouton de Gruyter. 355–499.
- Taylor D M (1977). *Languages of the West Indies*. Baltimore: John Hopkins University Press.
- Tovar A & De Tovar C L (1984). *Catálogo de las lenguas de América del Sur*. Madrid: Editorial Gredos.

Areal Linguistics

L Campbell, University of Utah, Salt Lake City, UT, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction: Defining the Concept

Areal linguistics is concerned with the diffusion of structural features across language boundaries within a geographical area. The term ‘linguistic area’ refers to a geographical area in which, due to borrowing and language contact, languages of a region come to share certain structural features – not just loan-words, but also shared phonological, morphological, syntactic, and other traits. The terms ‘sprachbund,’ ‘diffusion area,’ ‘adstratum relationship,’ and ‘convergence area’ are also sometimes used to refer to linguistic areas. The central feature of a linguistic area is the existence of structural similarities shared among languages of a geographical area, where usually some of the languages are genetically unrelated or at least are not all close relatives. It is assumed that the reason the languages of the area share these traits is because they are borrowed.

There are two sorts of linguistic area studies. The more common circumstantialist approach lists similarities found in the languages of a geographical area, allowing the list of traits to suggest diffusion, but typically without seeking the historical linguistic evidence which could demonstrate that the traits are indeed diffused. Circumstantialist areal linguistics has been criticized, since it does not eliminate chance, universals, and possibly undetected genetic relationships as alternative possible explanations for shared traits. The historicist approach seeks concrete evidence showing that the shared traits are diffused. The historicist approach is preferred because it is more rigorous and reliable, although the lack of clear evidence in many cases makes reliance on the circumstantialist approach necessary in some situations (Campbell, 1985).

Linguistic areas are often defined, surprisingly, by a rather small number of shared linguistic traits.

Examples of Linguistic Areas

A good way to get a solid feel for linguistic areas and how they are defined is to look at some of the better-known ones. In what follows, some of the best-known linguistic areas are inspected briefly together with the more important of the generally accepted defining traits shared by the languages of each area.

The Balkans

The Balkans is the best known of all linguistic areas. The languages of the Balkans are Greek, Albanian, Serbo-Croatian, Bulgarian, Macedonian, and Romanian (to which some scholars also add Romani and Turkish). Some salient traits of the Balkans linguistic area are the following:

1. a central vowel /i/ (or /ə/) (not present in Greek or Macedonian);
2. syncretism of dative and genitive cases (dative and genitive merged in form and function); this is illustrated by Romanian *fetei* ‘to the girl’ or ‘girl’s’, as in *am data o carte fetei* ‘I gave the letter to the girl’ and *frate fetei* ‘the girl’s brother’;
3. postposed articles (not in Greek), for example Bulgarian *mắt* ‘the man’ / *mă* ‘man’;
4. periphrastic future (future signaled by an auxiliary verb corresponding to ‘want’ or ‘have,’ not in Bulgarian or Macedonian), as in Romanian *voi fuma* ‘I will smoke’ (literally ‘I want smoke’) and *am a cînta* ‘I will sing’ (literally ‘I have sing’);
5. periphrastic perfect (with an auxiliary verb corresponding to ‘have’);
6. absence of infinitives (rather with constructions such as ‘I want that I go’ for ‘I want to go’);
7. double marking of animate objects by use of a pronoun copy, as in Romanian *i-am scris lui Ion* ‘I wrote to John’, literally ‘to.him-I wrote him John’, and Greek *ton vlépo ton Jáni* ‘I see John’, literally ‘him.ACC I see him.ACC John’ (Sandfeld, 1930; Schaller, 1975; Joseph, 1992).

South Asia (the Indian Subcontinent)

The South Asia linguistic area is composed of languages belonging to the Indo-Aryan, Dravidian, Munda, and Tibeto-Burman language families. Some traits shared among different languages of the area are the following:

1. retroflex consonants, particularly retroflex stops;
2. absence of prefixes (except in Munda);
3. presence of a dative-subject construction (that is, dative-experiencer, as in Hindi *mujhe maaluum tha* ‘I knew it’ [‘to me’ + ‘know’ + PAST]);
4. subject-object-verb (SOV) basic word order, including postpositions;
5. absence of a verb ‘to have’;
6. ‘conjunctive or absolutive participles’ – a tendency for subordinate clauses to have nonfinite verbs (that is, participles) and to be preposed; for example, relative clauses precede the nouns they modify;
7. morphological causatives;

8. so-called 'explicator compound verbs,' where a special auxiliary from a limited set is said to complete the sense of the immediately preceding main verb, and the two verbs together refer to a single event, as for example Hindi *le jaanaa* 'to take (away)' ('take' + 'go');
9. sound symbolic forms based on reduplication, often with *k* suffixed (for example in Kota, a Dravidian language: *kad-kadk* '[heart] beats fast with guilt or worry'; *a:nk-a:nk* 'to be very strong [of man, bullock], very beautiful [of woman]')

Some of these proposed areal features are not limited to the Indian subcontinent, but can be found also in neighboring languages (for example, SOV basic word order is found throughout much of Eurasia and northern Africa) and in languages in many other parts of the world. Some of the traits are not necessarily independent of one another; for example, languages with SOV basic word order tend also to have nonfinite (participial) subordinate clauses, especially relative clauses, and not to have prefixes (Emeneau, 1956; Masica, 1976; Emeneau, 1980; Emeneau, 2000).

Mesoamerica

The language families and isolates which make up the Mesoamerican linguistic area are Nahua (a branch of Uto-Aztecan), Totonacan, Otomanguean, Mixe-Zoquean, Mayan, Xinkan (Xinca), Tarascan (Purépecha), Cuitlatec, Tequistlatecan, and Huave. Five areal traits are shared by nearly all Mesoamerican languages, but not by neighboring languages beyond this area, and these are considered particularly diagnostic of the linguistic area:

1. nominal possession of the type *his-dog the man* 'the man's dog', as in Pipil (Uto-Aztecan) *i-pe:lu ne ta:kat*, literally 'his-dog the man';
2. relational nouns (locative expressions composed of noun roots and possessive pronominal affixes), of the form, for example, *my-head* for 'on me', as in Tzutujil (Mayan) *č-r-i:x* 'behind it, in back of it', composed of *č-* 'at, in', *r-* 'his/her/its' and *i:x* 'back', contrasted with *č-w-i:x* 'behind me', literally 'at-my-back';
3. vigesimal numeral systems, based on combinations of 20, such as that of Ch'ol (Mayan): *hun-k'al* '20' (1×20), *č-aʔ-k'al* '40' (2×20), *uf-k'al* '60' (3×20), *hoʔ-k'al* '100' (5×20), *hun-bahk* '400' (1×400), *č-aʔ-bahk* '800' (2×400), etc.;
4. nonverb-final basic word order (generally no SOV languages) – although Mesoamerica is surrounded by languages both to the north and south

which have SOV (subject-object-verb) word order, languages within the linguistic area have VOS, VSO, or SVO basic order;

5. a number of loan translation compounds (calques) shared by the Mesoamerican languages, including examples such as 'boa' = 'deer-snake,' 'egg' = 'bird-stone/bone,' 'lime' = 'stone(-ash),' 'knee' = 'leg-head,' and 'wrist' = 'hand-neck'.

Since these five traits are shared almost unanimously throughout the languages of Mesoamerica but are found almost not at all in the languages just beyond the borders of Mesoamerica, they are considered strong evidence in support of the validity of Mesoamerica as a linguistic area. Four of these five traits have essentially the same distribution, clustering at the borders of Mesoamerica. Such bundling is uncommon in linguistic areas.

A large number of other features are shared among several Mesoamerican languages, but are not found in all the languages of the area, while some other traits shared among the Mesoamerican languages are found also in languages beyond the borders of the area (for details see Campbell *et al.*, 1986).

The Northwest Coast of North America

The Northwest Coast, the best known North American linguistic area, includes Tlingit, Eyak, the Athabaskan languages of the region, Haida, Tsimshian, Wakashan, Chimakuan, Salishan, Alsea, Coosan (Coos), Kalapuyan (Kalapuya), Takelma, and Lower Chinook (Chinook). These languages are characterized by elaborate systems of consonants, which include series of glottalized stops and affricates, labiovelars, multiple laterals, and uvular stops in contrast to velars. There are typically few labial consonants (labials are completely lacking in Tlingit and Tillamook and are quite limited in Eyak and most Athabaskan languages); in contrast, the uvular series is especially rich in most of these languages. There are typically few vowels, only three (*i*, *a*, *o*, or *i*, *a*, *u*) in several of the languages, four in others. Several of the languages have pharyngeals (*ʕ*, *ħ*), and most have glottalized resonants and continuants. Shared morphological traits include: extensive use of suffixes; near absence of prefixes; reduplication (of several sorts, signaling iteration, continuative, progressive, plural, collective, distribution, repetition, diminutive, etc.); numeral classifiers; alienable/inalienable oppositions in nouns; evidential markers in the verb, and verbal locative-directional markers; masculine/feminine gender (shown in demonstratives and articles); visibility/invisibility opposition in demonstratives. Aspect is more important than tense. All but Tlingit have passivelike constructions. The

negative appears as the first element in a clause regardless of the usual word order. Overt marking of nominal plurals, as in many American Indian languages, is absent or limited. Northwest Coast languages also have lexically paired singular and plural verb stems (that is, a lexical root may be required with a plural subject which is entirely different from the root used with a singular subject).

Some other traits are shared by a smaller number of Northwest Coast languages (see Campbell, 1997: 333–334; cf. Sherzer, 1976).

The Baltic

The Baltic linguistic area includes at its core (Balto-) Finnic languages (especially Estonian and Livonian), Baltic languages (Indo-European), and Baltic German; however, all of the following have been included in different treatments of the Baltic linguistic area: Old Prussian (Prussian) (extinct), Lithuanian, and Latvian (Baltic languages); the ten Saami (Lapp) languages, Finnish, Estonian, Livonian, Votic (Votian), Vepsian (Veps), Karelian, and others (of the Finnic branch of Finno-Ugric); High German, Low German (Low Saxon), Baltic German, and Yiddish (Western Yiddish) (West Germanic); Danish, Swedish, and Norwegian (North Germanic); Russian, Belorussian, Ukrainian, Polish, and Kashubian (Slavic); Romani (Indo-Aryan, branch of Indo-European); and Karaim (Turkic).

Shared features of the Baltic area include the following:

1. first-syllable stress;
2. palatalization of consonants;
3. tonal contrasts;
4. partitive case/partitive constructions (to signal partially affected objects, equivalent to, for example, “I ate [some] apple”) in Finnic, Lithuanian, Latvian, Russian, Polish, etc.;
5. direct objects in the nominative case in a number of constructions which lack overt subjects (Finnic, Baltic, North Russian);
6. evidential mood: “John works hard (it is said/ reported/inferred)” (Estonian, Livonian, Latvian, Lithuanian);
7. prepositional verbs (as German *aus-gehen* [out-to-go] ‘to go out’): German, Livonian, Estonian, Baltic, and others;
8. subject-verb-object (SVO) basic word order;
9. agreement of adjectives in number with the nouns they modify (all languages of the area except Saami languages and Karaim); they also agree in case in all except the Scandinavian languages (which have lost case distinctions for adjectives);

they also agree in gender in Baltic, Slavic, and Scandinavian languages, as well as in German, Yiddish, and some others.

For a more complete list of traits attributed to the Baltic linguistic area, see Zeps, 1962; Koptjevskaja-Tamm, 2002; and especially Koptjevskaja-Tamm and Wälchli, 2001; compare also Jakobson, 1931.

Ethiopia

Languages of the Ethiopian linguistic area include: Beja (Bedawi), Awngi, Afar, Sidamo, Somali, etc. (Cushitic languages); Geez, Tigré, Tigrinya (Tigrigna), Amharic, etc. (Ethiopian Semitic languages); Wellamo (Wolaytta), Kefa (Kaficho), Janjero (Yemsa), etc. (Omotic languages); Anyuak (Anuak) and Gumuz (Nilo-Saharan languages); and others. Among the traits they share are the following:

1. SOV basic word order, including postpositions;
2. subordinate clause preceding main clause;
3. gerund (nonfinite verb in subordinate clauses, often inflected for person and gender);
4. a ‘quoting’ construction (a direct quotation followed by some form of ‘to say’);
5. compound verbs (consisting of a nounlike ‘pre-verb’ and a semantically empty auxiliary verb);
6. negative copula;
7. plurals of nouns not used after numbers;
8. gender distinction in second- and third-person pronouns;
9. reduplicated intensives;
10. a different present tense marker for main and subordinate clauses;
11. a form equivalent to the feminine singular used for plural concord (feminine singular adjective, verb, or pronoun is used to agree with a plural noun);
12. a singulative construction (the simplest noun may be a collective or plural and it requires an affix to make a singular);
13. shared phonological traits such as *f* but no *p*, palatalization, glottalized consonants, gemination, presence of pharyngeal fricatives (*ħ* and *ʕ*) (Ferguson, 1976; Thomason, 2001; cf. Tosco, 2000).

How Linguistic Areas Are Defined

The following criteria have at times been considered relevant for attempts to establish linguistic areas: (1) the number of traits shared by languages in a geographical area, (2) bundling of the traits in some significant way (for example, clustering at roughly

the same geographical boundaries), and (3) the weight or complexity of different areal traits (some are accorded more significance for determining areal affiliation on the assumption that they are more difficult to acquire than others).

To establish a linguistic area, the more shared features the better. Linguistic areas in which many diffused traits are shared among the languages are considered better established. Nevertheless, some scholars believe that even one shared trait is enough to define a weak linguistic area (Campbell, 1985). In any event, it is clear that some areas are more securely established than others because they are supported by more shared areal traits. In the linguistic areas described above, the number and kind of shared traits vary considerably.

The idea that greater weight or importance should be attributed to some traits for defining linguistic areas can be illustrated with the borrowed word order patterns in the Ethiopian linguistic area. Ethiopian Semitic languages exhibit a number of interconnected word order patterns which are borrowed from neighboring Cushitic languages. Several of these traits reflect the diffusion of the SOV basic word order typology of Cushitic languages into the formerly VSO Ethiopian Semitic languages. Typologically the orders noun-postposition, verb-auxiliary, relative clause-head noun, and adjective-noun are all correlated and tend to co-occur with SOV order cross-linguistically. Their presence in Ethiopian Semitic languages (some with all of these, others with somewhat fewer) might seem to reflect several different diffused traits (SOV counted as one, noun-postposition as another, and so on), and might be taken as several independent pieces of evidence supporting the existence of the linguistic area. However, from the perspective of expected word order co-occurrences, these word order arrangements are not independent traits, but reflect the diffusion of a single complex feature, the overall SOV word order type with its tendency for the various expected coordinated orderings in typologically interrelated constructions to co-occur. However, if borrowed SOV word order is counted as a single diffused areal trait, it must rank high in significance for defining a linguistic area, since it is much more difficult for a language to change so much of its basic structure under areal influence than it is to acquire less complex traits.

Some scholars had thought that the bundling of areal traits, clustering at the boundaries of a linguistic area, might be required for defining linguistic areas, though this has proven a poor criterion. Linguistic areas are similar to traditional dialects, where often one trait spreads across more territory than another

trait, so that their boundaries (or territories) do not coincide (do not 'bundle'). Typically the geographical extent of individual traits may vary considerably. However, in the rare situation where the traits do coincide at a clear boundary, the definition of a linguistic area matching their boundary is relatively secure. As mentioned, several of the traits in the Mesoamerican linguistic area do have the same boundary, but typically in other areas the areal traits do not share the same geographical boundaries, offering no clearly identifiable outer border of the linguistic areas in question.

Implications of Areal Linguistics for Linguistic Reconstruction and Subgrouping

Areal diffusion can have important implications for comparative reconstruction and for subgrouping within known language families. Nootkan provides a good example which illustrates this. The sound correspondences upon which Nootkan subgrouping is based are given in Table 1.

Nitinat and Makah appear to share the innovation which changed nasals to corresponding voiced stops (in [1]–[4]), while Nitinat and Nootka appear to share the change of the glottalized uvulars to pharyngeals (in [5] and [6]). (Makah and Nitinat also share the retention of uvular fricatives, which Nootka has changed to a pharyngeal [in (7) and (8)]; however, shared retentions are not valid evidence for subgrouping.) Here, one innovation (denasalization) suggests a subgrouping of Makah and Nitinat together, with Nootka more distantly related, while the other innovation (pharyngealization) suggests Nitinat and Nootka together, with Makah less closely related. This seeming impasse is solved when we take into account the fact that the absence of nasals is an areal feature shared by several other languages of the area; it diffused into both Makah and Nitinat

Table 1 Nootkan sound correspondences

| | <i>Makah</i> | <i>Nitinat</i> | <i>Nootka</i> | <i>Proto-Nootkan</i> |
|----|-----------------|----------------|---------------|----------------------|
| 1. | b | b | m | *m |
| 2. | b' | b' | m' | *m' |
| 3. | d | d | n | *n |
| 4. | d' | d' | n' | *n' |
| 5. | q' | ʕ | ʕ | *q' |
| 6. | q' ^w | ʕ ^w | ʕ | *q' ^w |
| 7. | χ ^w | χ ^w | ħ | *χ ^w |
| 8. | χ | χ | ħ | *χ |

(Haas, 1969).

under areal influence and is thus not real evidence of a shared common development before the languages separated; rather, it reached these two languages independently from elsewhere in the linguistic area. The innovation shared by Nitinat and Nootka of glottalized uvulars changing to pharyngeals (in [5] and [6]) is real evidence of subgrouping – a true (nondiffused) shared innovation. So, Nitinat and Nootka together constitute one branch of the family, Makah the other branch. Moreover, with respect to areal implications for reconstruction, if we did not know about the areal diffusion in this case, we might be tempted to reconstruct the voiced stops in Proto-Nootkan, since they occur in more languages than the nasals do, and to postulate a change of these to nasals in Nootka (for [1]–[4]), getting it wrong in this case. Thus, areal linguistic traits can have important implications for classification (subgrouping) and for reconstruction.

Areal Linguistics and Proposals of Distant Genetic Relationship

Some similarities among languages which are due to areal diffusion are often mistakenly taken to be evidence of a possible distant family relationship among languages whose classification is in question. The Mosan hypothesis, which proposes a genetic relationship between the Salishan, Wakashan, and Chimakuan language families, illustrates this problem, which is common in many instances of long-range comparison. Several scholars noted structural similarities among these Northwest Coast languages, but the Mosan hypothesis was not found convincing because much of the evidence turned out to rely on areal traits widely borrowed in the Northwest Coast linguistic area. Swadesh (1953) presented 16 shared structural similarities in support of Mosan, but most of these are Northwest Coast areal features (some of the traits are also typologically commonplace, found independently in languages throughout the world), for example:

1. “Extensive use of suffixes.”
2. “Nearly complete absence of functioning prefixes in Chimakuan and Wakashan, minor role in comparison to the suffixes in Salish.” (Typologically it is not unusual for suffixing languages to lack prefixes.)
3. “Extensive use of stem reduplication, including initial reduplication . . . and . . . full stem reduplication.”
4. “Aspect, including at least the dichotomy of momentaneous and durative.”
5. “Tense is an optional category.”

6. “Plural is an optional category.”
7. “Dichotomy of non-feminine versus feminine gender shown in demonstratives and articles.”
8. “Numeral classifier notions, shown by suffixes.”
9. “Two alternate stems for number” (lexically paired distinct singular and plural verb stems).
10. “Lexical suffixes . . . referring to body parts and other space references.”
11. “Predicative use of nouns.”
12. “Demonstrative distinctions such as the present versus absent, or visible versus invisible.”

As is clear, the traits which Swadesh listed as evidence for the Mosan hypothesis are better explained as the results of diffusion within the Northwest Coast linguistic area (see Campbell, 1997 for details.)

From this case, it is easy to see why the identification of areal traits is so important in historical linguistics. In this instance, failure to recognize the areal borrowings led to an erroneous proposal of genetic relationship among neighboring language families.

Kinds of ‘Linguistic Area’

It is generally recognized that things that have been called linguistic areas include entities with widely divergent character and historical backgrounds, depending on the social, cultural, political, geographical, attitudinal, and other factors which correlate with diffusion of linguistic features in different regions (Dahl, 2001: 1458; Kuteva, 1998: 308–309). As Thomason (2001: 104) explained,

[linguistic areas] arise in any of several ways – through social networks established by such interactions as trade and exogamy, through the shift by indigenous peoples in a region to the language(s) of invaders, through repeated instances of movement by small groups to different places within the area.

One finds in the literature many different sorts of linguistic areas, such as: incipient ones, only beginning to form and with as yet few shared traits; moribund and decaying ones, where, due to many changes after the area was actively formed, fewer traits are currently recognizable among the languages; overlapping ones, where different areas formed on top of or partially overlapping one another at different times for different reasons; multilateral (areal traits spreading from various languages of the region) versus unilateral areas (with the traits shared throughout the languages of an area stemming predominantly from one language); areas due to rapid conquest, population spread, and migration (traits moving with movement of speakers), others through home-grown, stay-in-place contact (movement of traits but not of peoples); and disrupted areas with

“latecomers, earlier drop-outs, and temporary passers-by” (Stolz, 2002: 265). “In short, the notion ‘linguistic area’ does not refer to a uniform phenomenon, either socially or linguistically” (Thomason, 2001: 115). This array of different kinds of linguistic area raises questions about whether the notion of ‘linguistic area’ is warranted, whether all these different ‘objects’ legitimately qualify as ‘linguistic areas,’ given their very different natures and composition and given the very different circumstances of their birth (and decay). The notion of ‘linguistic area’ offers little upon which these different sorts of linguistic areas can be united other than the fact that they all involve borrowing in some way, but borrowings of different sorts, for different reasons, in different settings, and at different times.

Linguistic Areas versus Borrowing Generally

It is generally acknowledged that linguistic areas are “notoriously messy,” “notoriously fuzzy” things (Thomason and Kaufman, 1988: 95; Heine and Kuteva, 2001: 396; Tosco, 2000: 332), and that “what we understand about linguistic areas is depressingly meager” (Thomason, 2001: 99). A common perception is that the term ‘linguistic area’ is difficult to define (cf. Heine and Kuteva, 2001: 409). As Thomason (2001: 99) observed, “linguistics has struggled to define the concept ever since [Trubetzkoy, 1928], mainly because it isn’t always easy to decide whether a particular region constitutes a linguistic area or not.” Stolz (2002: 259) believed that “the search for clearcut definitions [of ‘Sprachbund, linguistics area, and areal type’] has been largely futile and will probably never come to a really satisfying conclusion.” In spite of prolonged efforts to define ‘linguistic area,’ there is no general agreement on its definition, and even for the most widely accepted linguistic areas, such as the Balkans, scholars do not agree wholly on which languages belong to the area, what linguistic traits characterize the area, and what its precise geographical extent is. This difficulty has been related to the lack of clear distinction between areal phenomena and borrowing generally (Campbell, in press). Thus Dahl (2001: 1458) asked:

In the end, we are led to the following more far-going question about the notion of area: to what extent do areas ... have a reality of their own and to what extent are they just convenient ways of summarizing certain phenomena? At the most basic level, linguistic contact relationships are binary: one language influences another. An area is then simply the sum of many such binary relationships.

Campbell (in press) argues that the various definitions of ‘linguistic area’ offered in the literature confirm that linguistic areas amount to just the study of local linguistic borrowing and its history. Every ‘linguistic area,’ to the extent that the notion has any meaning at all, arises from an accumulation of individual cases of ‘localized diffusion’; it is the investigation of these specific instances of diffusion, and not the pursuit of defining properties for linguistic areas, that increases our understanding and explains the historical facts. With the focus rather on specific instances of borrowing, many of the unresolved issues and indeterminacies which have dogged areal linguistics from the outset cease to be relevant questions. It is the diffused linguistic changes themselves that count and not the attempt to seek meaning in the geography that secondarily is involved (Campbell, 2004). A linguistic area, to the extent that it may have a legitimate existence at all, is merely the sum of borrowings in individual languages in contact situations. If we focus rather on understanding borrowings, those contingent historical events, the difficulty of determining what qualifies as a legitimate linguistic area ceases to be a problem.

Bibliography

- Bright W (ed.) (1992). *International encyclopedia of Linguistics* (4 vols). Oxford: Oxford University Press.
- Campbell L (1985). ‘Areal linguistics and its implications for historical linguistic theory.’ In Fisiak J (ed.) *Proceedings of the Sixth International Conference of Historical Linguistics*. Amsterdam: John Benjamins. 25–56.
- Campbell L (1997). *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press.
- Campbell L (2004). *Historical linguistics: an introduction* (2nd edn.). Edinburgh: Edinburgh University Press/Cambridge, MA: MIT Press.
- Campbell L (in press). ‘Areal linguistics: the problem to the answer.’ In Matras Y, McMahon A & Vincent N (eds.) *Language contact and areal linguistics*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Campbell L, Kaufman T & Smith-Stark T (1986). ‘Meso-america as a linguistic area.’ *Language* 62, 530–570.
- Dahl Ö (2001). ‘Principles of areal typology.’ In Haspelmath M, König E, Oesterreicher W & Raible W (eds.) *Language typology and language universals: an international handbook* 2. Berlin: Mouton de Gruyter. 1456–1470.
- Emeneau M B (1956). ‘India as a linguistic area.’ *Language* 32, 3–16.
- Emeneau M B (1980). *Language and linguistic area: essays by Murray B Emeneau*, selected and introduced by A S Dil. Palo Alto, CA: Stanford University Press.

- Emeneau M B (2000). 'Linguistics in India: past and future.' In Lockwood D G, Fries P H & Copeland J E (eds.) *Functional approaches to language, culture, and cognition: papers in honor of Sydney B Lamb*. Amsterdam: John Benjamins. 545–554.
- Ferguson C (1976). 'The Ethiopian language area.' In Bender M L, Bowen J D, Cooper R L & Ferguson C A (eds.) *Language in Ethiopia*. Oxford: Oxford University Press. 63–76.
- Haas M R (1969). *The prehistory of languages*. The Hague: Mouton.
- Heine B & Kuteva T (2001). 'Convergence and divergence in the development of African languages.' In Aikhenvald A Y & Dixon R M W (eds.) *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press. 393–411.
- Jakobson R (1931). 'Über die phonologischen Sprachbünde.' *Travaux du Cercle Linguistique de Prague* 4, 234–240.
- Joseph B (1992). 'Balkan languages.' In Bright (ed.) 153–155.
- Koptjevskaja-Tamm M (2002). 'The Circum-Baltic languages: a contact-superposition zone in the European periphery.' In Ramat & Stolz (eds.) 209–222.
- Koptjevskaja-Tamm M & Wälchli B (2001). 'The Circum-Baltic languages: an areal-typological approach.' In Dahl Ö & Koptjevskaja-Tamm M (eds.) *Circum-Baltic languages*. Amsterdam: John Benjamins. 615–750.
- Kuteva T (1998). 'Large linguistic areas in grammaticalization: auxiliaries in Europe.' *Language Sciences* 20, 289–311.
- Masica C P (1976). *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.
- Masica C P (1992). 'Areal linguistics.' In Bright (ed.) 108–112.
- Ramat P & Stolz T (eds.) (2002). *Mediterranean languages: papers from the MEDTYP workshop, Tirrenia, June 2000*. Bochum: Brockmeyer.
- Sandfeld K (1930). *Linguistique balkanique: problèmes et résultats*. Paris: Champion.
- Schaller H W (1975). *Die Balkansprachen: eine Einführung in die Balkanphilologie*. Heidelberg: Winter.
- Sherzer J (1976). *An areal-typological study of American Indian languages north of Mexico*. Amsterdam: North Holland.
- Stolz T (2002). 'No sprachbund beyond this line! On the age-old discussion of how to define a linguistic area.' In Ramat & Stolz (eds.) 259–281.
- Swadesh M (1953). 'Mosan I: a problem of remote common origin.' *International Journal of American Linguistics* 19, 26–44.
- Thomason S G (2001). *Language contact: an introduction*. Edinburgh: Edinburgh University Press.
- Thomason S G & Kaufman T (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tosco M (2000). 'Is there an Ethiopian language area?' *Anthropological Linguistics* 42, 329–365.
- Trubetzkoy N S (1928). Proposition 16. *Acts of the First International Congress of Linguists*, 17–18. Leiden.
- Zeps V (1962). *Latvian and Finnic linguistic convergences*. Bloomington: Indiana University Press.

Armenian

B Vaux, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

© 2006 Elsevier Ltd. All rights reserved.

'Armenian' actually refers to several languages, including Standard Eastern and Western Armenian, Middle (/Medieval/Cilician) Armenian, and Classical Armenian, as well as Zok, formerly spoken by the Armenian inhabitants of southeastern Nakhichevan; Kistinək, spoken by the Armenian inhabitants of Musaler, Turkey; Kesbənuək, spoken by the Armenian inhabitants of Kesab, Syria; Homshetsma or Homshetsnak (referred to as Hemşince in Turkish), spoken by the Hemshinli of northeast Turkey and the Hamshen Armenians of the Black Sea coastal regions of Abkhazia and Russia; and dozens of other mutually unintelligible variants of Armenian originally spoken in Turkey, Armenia, Azerbaijan, Iran, Georgia, Abkhazia, Russia, and Israel. Lomavren, the language of the Boshia (or Posha) gypsies of Turkey

and Armenia, draws its grammar from the Erzerum dialect of Armenian but its lexicon is mostly of Indic origin; it therefore is not clear whether or not the language should be classified as a form of Armenian. All employ the Armenian alphabet (created by Mesrob at the beginning of the 5th century) except for the Turkish forms of Homshetsma, which normally appear only in oral contexts, but in recent years have begun to show up in Turkish orthography in collections of word lists from minority groups in Turkey, lyrics on CDs, and the like.

Armenian belongs to the Indo-European family, and is commonly believed to be most closely related to Greek and Indo-Iranian. (For instance, all three share a prohibitive particle **me*: (Greek *me*-, Sanskrit *ma*-, Armenian *mi*) and the imperfect third-person singular augment **e*- (as in Greek *e-pher-e*, Sanskrit *a-bhar-a-t*, Armenian *e-ber* '(s)he/it carried'). Many more such parallels are discussed in Clackson, 1994.) Because of its many loans from various Middle Iranian languages, especially Parthian, Armenian

was thought to be an Iranian dialect until Heinrich Hübschmann demonstrated in 1875 that it was a distinct branch of the Indo-European family. Scholars disagree on how the Armenians came to historical Armenia, the eastern half of present-day Turkey centered around Lake Van and Mount Ararat; some believe they came southward from the Russian steppe, others believe they and the Hittites came eastward from Greece, and others suggest they moved only a short distance from an original Indo-European homeland in the Transcaucasus. It is most likely that this settlement occurred in the second millennium B.C. The earliest mentions of the Armenians occur in the inscriptions of the Achaemenid Persian king Darius (6th century B.C.) and the Greek historian Herodotus (5th century B.C.)

The earliest written records of the Armenian language date from the 5th century A.D. shortly after the conversion of the Armenians to Christianity in the 4th century led to the creation of an Armenian alphabet by Mesrob around 401 and a systematic program of translating the books of the Bible. The language of the earliest translations was Classical Armenian (also called *grabar*, ‘written [language]’), which continued as the preferred literary form of Armenian until the 19th century, when it was supplanted by the three modern literary dialects.

In linguistic terms Armenian is notable for its significant divergences from Proto-Indo-European, particularly in terms of pronunciation and vocabulary. Some of the more striking phonological changes are the development of a rich set of affricates (*ts*, *dz*, etc.), the loss of final syllable rimes (e.g., PIE **worgʷom* ‘work’ > Classical Armenian *gorts*), the change of initial **dw* to *erk-* (e.g., PIE **dwo*: ‘2’ > Classical Armenian *erku*), and the change of original **w* to *g*. Most striking in the vocabulary of Armenian is the rarity of words inherited from Indo-European and the overwhelming predominance of words of unknown origin. Unsurprisingly, native IE words survive primarily in the core vocabulary: *mayr* ‘mother’ < **ma:ter*, *hayr* ‘father’ < **pater*, *kʰoyr* ‘sister’ < **swe-sor*, *kov* ‘cow’ < **gwows*, *tun* ‘house’ < **domos*, *em* ‘I am’ < **esmi*. The remainder of the lexicon is drawn primarily from Parthian, and to a lesser extent Greek and Syriac (q.v. Hübschmann, 1895); several hundred and perhaps as many as several thousand words are of unknown origin, most likely having come from Urartian, Hurrian, and other now-extinct autochthonous languages. Armenian also incorporated large numbers of Arabic words following the expansion of the Arabs in the Middle East in the 7th century, and the spoken language absorbed thousands of Turkish words following the arrival of Turkic tribes in Anatolia beginning in the 11th century.

Though there are dozens of mutually unintelligible varieties of Armenian, all share certain features. Proto-Armenian had four verbal conjugations, characterized by theme vowels -e-, -i-, -a-, and -u- (*ber-e-m* ‘I carry’, *χawsim* ‘I speak’, *χndam* ‘I rejoice’, *zgen-u-m* ‘I wear’); most modern dialects (including the Western and Eastern literary languages) have completely or partially lost the -u- conjugation, and standard Eastern Armenian has merged the -i- conjugation into the -e- conjugation. There were originally three morphologically distinct sets of personal endings for verbs – present, imperfect, and aorist – which were used in combination with additional tense and aspect markers to form the various tenses and moods. The system of nominal morphology in Proto- and Classical Armenian was rich, preserving the IE nominative, accusative, genitive, dative, instrumental, ablative, and locative cases in both singular and plural (but the IE dual was lost); there were at least eight different declensions, distinguished primarily by different theme vowels. This system was significantly reduced by the medieval period; Middle Armenian and the modern varieties now use the singular endings for the plural as well, and have only one productive declension, formed from parts of the original -i- and -o- declensions. With the exception of pronouns, the inventory of cases has significantly reduced as well: the accusative has merged with the nominative, and the genitive with the dative. Proto-Armenian had several participial forms, but only two of these survive into the modern period: the original past participle -eal is now -el in the Eastern dialects, and the original present participle -oꝛ is now used as a present participle and for relativizing subjects of subordinate clauses, as in the following Standard Western Armenian example:

| | | | |
|--------------------------------------|-----------|---------------|------------|
| qjn | kʰirkʰ-ə | kʰən-oꝛ | gin-ə |
| that | book-def. | buy-pres.ppl. | woman-def. |
| ‘the woman that is buying that book’ | | | |

The Western dialects have replaced -eal with -ats (> -adz) for past participles; all modern dialects also use the -ats participle to relative non-subjects of subordinate clauses, as in the following Western example:

| | | |
|-----------------------------------|---------------|--------------------------|
| (kʰu) | kʰən-adz | kʰirkʰ-ətʰ |
| 2sg _{GEN} | buy-past.ppl. | book-2sg _{POSS} |
| ‘the book that you (have) bought’ | | |

Most of the changes between Classical and Modern Armenian first appear in the medieval period in Middle Armenian documents, associated with the Armenian kingdom of Cilicia, which flourished from the 11th to 15th centuries A.D. in what is now south-central Turkey. Middle Armenian is generally

Western in character, though it shares many features with Eastern dialects as well. It inverts the pronunciation of the Classical Armenian plain voiced and voiceless stops (e.g., *berem* ‘I carry’ > *perem*, *pat* ‘wall’ > *bad*), a feature that is preserved in the modern Cilician dialects of Zeytun and Hadjin but differs from the Western and Eastern literary varieties (Eastern preserves the Classical system [bɛrɛm]); Western devoices and aspirates the original voiced series [p^hɛrɛm]). The Cilician kingdom was in close contact with several Crusader kingdoms; as a result, it borrowed a significant number of words from Crusader French, most famously what comes out as the standard Western form for ‘mister’, *baron*.

In the 19th century Armenian nationalists became interested in developing a literary form of the modern language. This was brought about by excising most Turkish forms from the regional dialects and replacing them with new borrowings from the classical language. The intellectual center around which the new Western literary language was organized was Constantinople (modern Istanbul), though many features of the standard dialect (including the pronunciation of the consonants) do not come from the Armenian dialect originally spoken there. The same holds for Eastern Armenian with respect to Erevan. The relationship between the two modern literary dialects is somewhat complicated; there are many grammatical differences (e.g., *W gə sirəm* vs. *E sirum em* ‘I love’, *W bidi sirəm* vs. *E kəsirəm* ‘I will love’ (note that the same form is used for the present in *W* and the future in *E*) and lexical differences (e.g., *W dʒɛrmag* vs. *E spitak* ‘white’; *W həs* vs. *E ɛstɛk* ‘here’, *W bɛdk^haran* vs. *E zuk^haran* ‘bathroom’, *W havgit^h* vs. *E dzu* ‘egg’), and most Western speakers have difficulty understanding Eastern, but many Eastern speakers are relatively comfortable with the Western dialect. This asymmetry in mutual intelligibility most likely results from the fact that large numbers of speakers of Western dialects fled to Eastern Armenia following the Russo-Turkish war in 1828 and the Turkish Genocide in 1915–1920, whereas before the fall of the Soviet Union in 1991 most Western Armenians had little or no exposure to Eastern Armenian. The fact that there is some mutual intelligibility in both directions can also be linked to the fact that the literary dialects tend to borrow the same forms from Classical Armenian, and (at least in recent decades) employ the same newly coined words.

The destruction of the Armenian homeland and more than a million Armenians by the Ottoman government in 1915–1920 rendered most nonstandard varieties of modern Armenian moribund; with few exceptions the Armenians in the diaspora (primarily Lebanon, France, and notably in the Los Angeles area

Table 1 The Armenian alphabet, with IPA equivalents for eastern pronunciation

| | | | | | |
|---|----------------|---|-----------------|----|-----------------|
| ա | a | ծ | ts | ջ | dʒ |
| բ | b | կ | k | ռ | r |
| գ | g | հ | h | ս | s |
| դ | d | ձ | dz | վ | v |
| ե | (j)ɛ | ղ | ɣ | տ | t |
| զ | z | ճ | tʃ | ր | ɾ |
| է | ɛ | մ | m | ց | ts ^h |
| ը | ə | յ | j | ու | u |
| թ | t ^h | ն | n | փ | p ^h |
| ժ | ʒ | շ | ʃ | ք | k ^h |
| ի | i | ռ | (v)o | օ | ɔ |
| լ | l | չ | tʃ ^h | ֆ | f |
| խ | χ | պ | p | | |

of the United States) speak only Standard Western Armenian. There were approximately 6.8 million speakers of Armenian in 1996, but all varieties of the language except for Standard Eastern Armenian are in immediate danger of extinction as very few diaspora Armenians under the age of 30 speak the language fluently.

Whereas Classical Armenian was relatively Indo-European in its syntactic and morphological structure, all varieties of Modern Armenian are typologically much closer to Turkish and the Balkan languages. Compare, for instance, the formation of relative clauses, exemplified by ‘I saw the bird that was singing in the tree’: Classical – *tesɪ əz-t^hərtʃ^hun-ən* or *etɛr i veraj tsar-ɔj-n* (I.saw specific-bird-definite that was.-singing in on tree-genitive-definite), Western – *dzar-i-n vərə jɛrk^hɔz t^hərtʃ^hun-ə dəsa* (tree-gen.-def. on singing bird-def.I.saw). Western Armenian has undergone additional influence from Turkish and Greek (cf. *sɛpɛɾin* ‘carrot’, *istak^hɔz* ‘lobster’, *bant^hɔz* ‘hotel’), whereas Eastern Armenian has been heavily influenced by Russian (e.g., the standard form for ‘potatoes’ is *k^hart^hɔfli*, and the word for ‘gay’ is *galubɔj*, from the Russian word originally meaning ‘sky blue’; the native word for ‘blue’, *kapujt*, cannot be used in this sense).

The Lord’s Prayer in Different Varieties of Armenian, Rendered in the IPA Classical Armenian (Edžmiatsin ms. 229, 989 A.D.)

Հայր մեր որ յերկինս. սուրբ եղիցի անուն քո.
եկեցէ արքայութիւն քո. եղիցին կամք քո որպէս
յերկինս եւ յերկրի. զհաց մեր հանապազորդ
սուր մեղ պատար. եւ թող մեզ զպարտիս մեր.
որպէս եւ մեք թողումք մերոց պարտապանաց.
եւ մի տանի որ զմեզ ի փորձութիւն. այլ՝ փրկեա
զմեզ ի չարէ. զի քո է արքայութիւն եւ զաւրութիւն
եւ փառք յաւիտեանս ամէն:

hajr mer or jerkinəs. surb ɛkɪts^hi anun k^ho. ekets^hɛ
ark^hajut^hiwn k^ho. ɛkɪts^hin kamk^h k^ho orpɛs
jerkinəs ew jerkri. əzhats^h mer hanapazord tur mez
ajɟawɾ. ew t^hoɾ mez əzpartis mer. orpɛs ew mek^h
t^hoɾumk^h merots^h partapanats^h. ew mi tanir əzmez
i p^hordzut^hiwn. ajɟ^h p^hərkea əzmez i t^harɛ. zi k^ho ɛ
ark^hajut^hiwn ew zawrut^hiwn ew p^hark^h jawiteanəs.
amen.

Standard Eastern Armenian

Հայր մեր, վոր յերկնքում ես. սուրբ թող լինի
քո անունը. քո թագաւորութիւնը թող գա. քո
կամքը թող լինի յերկրի վրա, ինչպէս վոր յերկնքում ե.
մեր հանապազօրյա հացը տուր մեզ
այսօր. յեւ թող մեզ մեր պարտքերը, ինչպէս յեւ
մենք ենք թողնում մեր պարտականներին. յեւ
մի տար մեզ փորձութեան, այլ փրկիր մեզ չարից.
վորովհետեւ քօնն ե թագաւորութիւնը յեւ
զորութիւնը յեւ փառքը համլիտյանս. ամէն:

hajr mer, vor jerkənk^hum ɛs. surp^h t^har lini k^ho
anunə. k^ho t^hagavərut^hjuna t^har ga. k^ho kamk^hə
t^har lini jerkri vərə, intj^hpɛs vor jerkənk^hum ɛ. mer
hanapazordja hats^hə tur mez ajɟə. jev t^har mez mer
partk^hɛrə, intj^hpɛs jev menk^h enk^h t^harɛnum mer
partakanɛrin. jev mi tar mez p^hordzut^hjan, ajl
p^harkir mez t^harits^h. vorəvhetev k^hon: ɛ
t^hagavərut^hjuna jev zərut^hjuna jev p^hark^hə
havitjanəs. amen.

Standard Western Armenian

Ով հայր մեր որ երկինքն ես, քու անունդ սուրբ
ըլլա. քու թագաւորութիւնդ գա. քու կամքդ ըլլա
ինչպէս երկինքը՝ նոյնպէս երկրի վրայ. մեր
ամէն օրուան հացը այսօր ալ մեզի տուր, մեզի ներք
մեր պարտքերը ինչպէս մենք ալ կը ներենք մեր
պարտականներուն. ու մեզ փորձութեան մի
տանիր, հապա չարէն մեզ ազատէ. քանզի քուկդ
է թագաւորութիւնը եւ զորութիւնը ու փառքը
յաւիտեանս: Ամէն:

ov hajr mer vor jergink^hn ɛs, k^hu anunəth surp^h əl:a.
k^hu t^hak^havərut^hjuna t^hə. k^hu gamk^həth əl:a
intj^hbɛs jergink^hə, nujnbɛs jergri vərə. mer amen
orvan hats^hə ajɟə al mezi dur, mezi nɛrɛ mer
bardk^hɛrə intj^hbɛs menk^h al gə nɛrɛnk^h mer
bardaganɛrun. u mez p^harts^hut^hjan mi danir, haba
tj^harɛn mez azadə. k^hanzi k^hugəth ɛ
t^hak^havərut^hjuna jev zərut^hjuna u p^hark^hə
havidjanəs. amen.

Zeytun Dialect (Cilicia, South-Central Turkey)

ov mej bobə vj ijgink^hn-is, k^hu anunəth surp t^harəna.
k^hu t^hek^hevvyt^hynəth t^har kə. k^hu gəmk^həth t^har la,
intj^hbɛs ijgink^hə, inden el ijgejin vijo. mij amen
əjven həts^hə ɛsəj miz tuj. jev miz nɛjɛ mij bəjɟk^hə,
tj^həts^h vər mink^h el gə nɛjink^h mij bəjɟk^hi dəjɛrun.
jev miz p^həjtsut^han mi danəj, habə tj^həjen miz
azadə. tj^hunk^hi k^hin: ɛ t^hek^hevvyt^hynəth jev
zəjut^hynə u p^hark^hə. havidjanəs havidənits^h. amen.

Kesab

əv mier bybə, surp ɛkni k^hɛ ænun,
k^het^hek^hevvyrut^hynə t^har kə, k^hɛ iradəthəth ən:ə,
tj^hytshər k^hi igrənk^hə t^harzen el ikədinə, mier amen
evyr hoəts^hə dur miez ɛs evyr el, mier bərdk^hə miezi
bəətj^hlamuf əvə tj^hytshər k^hi mienk^h ginonk^h
mierontshə, vɛ zəzmiez p^hərtsyt^hjan mi danə, habə
xələsə i tj^harien, tj^hynk^hi k^hɛ ɛ t^hek^hevvyrut^hynə,
tj^hɛrɛfə, k^huvet^hə, havidicinəs havidənits^h amən.

Bibliography

- Adjarian H (1909). *Classification des dialects arméniens*. Paris: Champion.
- Bardakjian K & Vaux B (1998). *Eastern Armenian: a text-book*. Ann Arbor: Caravan.
- Clackson J (1994). *The linguistic relationship between Armenian and Greek*. Oxford: Blackwell.
- Dumézil G (1964). *Notes sur le parler d'un Arménien musulman de Hemşin*. Brussels: Palais des Académies.
- Finck F N (1907). *Die Sprache der armenischen Zigeuner*. St. Petersburg.
- Hübschmann H (1895). *Armenische Grammatik: I. Theil Armenische Etymologie*. Leipzig.
- Karst J (1901). *Historische Grammatik des Kilikisch-Armenischen*. Strasbourg: Trübner.
- Meillet A (1936). *Esquisse d'une grammaire comparée de l'arménien classique* (2nd edn.). Vienna: Imprimerie des Pères mékhitharistes.
- Thomson R (1975). *An introduction to classical Armenian*. Delmar, NY: Caravan.
- Vaux B (1998). *The phonology of Armenian*. Oxford: Oxford University Press.

Arrernte

G Breen, Institute for Aboriginal Development, Alice Springs, NT, Australia

© 2006 Elsevier Ltd. All rights reserved.

Arrernte, using the most common current spelling, is the name of what was for many years, if it is not still, the most well-known Aboriginal language name in Australia. The first widely used spelling was 'Arunta,' and this is the spelling that leads to the best approximation to the pronunciation of the name ([ɔ̃ʀəŋdə], often [ʀəŋdə]) by the general English-speaking reader. The early German Lutheran missionaries introduced the spelling 'Aranda.' The spelling 'Arrernte' is that of the practical orthography most

used now by writers in the language, and has acquired wide currency, for example, in the print media. The (Lutheran) Finke River Mission now uses the spelling 'Arrarnta.'

At the time of European settlement, which reached the central part of Australia in the 1870s, the Arrernte speakers occupied a large area in the southeastern part of the present Northern Territory, spilling over into Queensland and South Australia (see Figure 1). The name Arrernte (with various qualifiers) is used for several dialects of what is generally regarded as a single language, called Upper Aranda by Hale (1962), and also for another closely related language that Hale called Lower Aranda. The Upper Aranda language group includes three main subgroupings: Western,

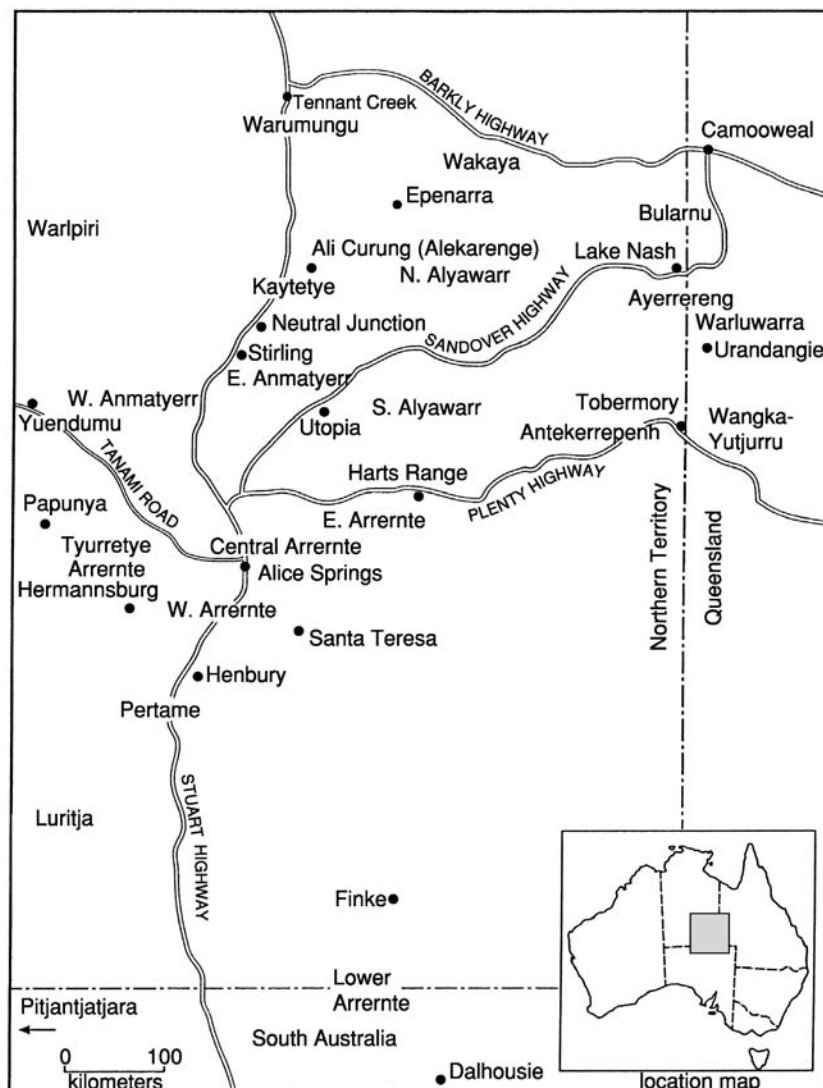


Figure 1 Locations in Australia where Arrernte and some neighboring languages are spoken. Reprinted from Green (1998), Kin and country: aspects of the use of kinterms in Arandic Languages. M.A. thesis, University of Melbourne.

Central, and Eastern Arrernte and Anmatyerr and Alyawarr as dialects, which have of the order of a 1000 speakers each and are still being learned by children; Southern Arrernte (or Pertame), many of whose speakers now use Western Arrernte; and extinct or nearly extinct varieties, Antekerrepnh and Ayerrereng. The relationship of modern Western Arrernte to an almost extinct dialect that has been called Tyurretye Arrernte (Breen, 2001) is not clear; it could be that the latter is the original western dialect and the former is essentially Southern Arrernte. Anmatyerr and Alyawarr do not identify as Arrernte. The other language, Lower Arrernte (Hale's Lower Aranda, called Lower Southern Arrernte/Aranda by some, but Arrernt Imarnt in the dictionary that is being compiled by the author at the time of writing), had in 2004 only a couple of moderately competent, elderly speakers. The two languages, Upper and Lower Arrernte, are grouped (but not uncontroversially) with the more distantly related Kaytetye under the name Arandic (see Koch, 2004), and this is classed as a subgroup of the Pama–Nyungan family (which, again, is not universally accepted).

Study of the Arrernte language was begun after 1877, by German missionaries, notably Carl Strehlow. Strehlow's son, T. G. H. Strehlow, continued his father's study of the language and amassed a vast quantity of data, the culmination of his work being the wonderful *Songs of Central Australia* (1971). Somewhat earlier, around 1960, the linguist Ken Hale (1934–2001) had collected excellent material in most dialects. The Summer Institute of Linguistics (SIL) and other mission linguists have also worked on the languages for many years; the first substantial Western Arrernte Bible portion appeared in 1925 and there have been many other, mostly smaller, works. Substantial theses on Arrernte phonetics and grammar have been written by David Wilkins (1989), John Henderson (1998), and Victoria Anderson (2000). One of the most extensive dictionaries of any Australian language to appear to date is that of Eastern and Central Arrernte by Henderson and Dobson (1994).

Smaller dictionaries are those of Alyawarr (Green, 1992) and Western Arrernte (Breen, 2000); dictionaries of Anmatyerr (a work in progress, by Jenny Green) and Lower Arrernte (Breen) are to appear in the near future. No detailed grammar has been published. Indigenous writing is in its infancy.

Table 1 gives the consonant inventory of Central Arrernte, using orthographic symbols, as typical of these languages. The basic vowel system comprises a featureless vowel, written *e*, dependent for its quality on the surrounding consonants, and a low vowel, *a*. In most dialects there is also a high front vowel, with a comparatively small functional load, and in some there may also be a high back vowel, with a small functional load, which, however, may be better analyzed as due to the effect on *e* of roundness on a following consonant. Roundness, derived from an ancestral rounded vowel, may be associated with consonant positions. A seventh consonant position, prepalatalized apical (*yt*, *yn*, *ytñ*, *yl*), postulated for some dialects, may be more correctly analyzed as a palatalization feature associated with the consonant position. In other dialects, prepalatalized apicals are an allophone of phonemes in the series called apical postalveolar.

Arrernte (and also Kaytetye) are a focus of attention for linguists because of the substantial sound changes that the languages have undergone in the not too distant past. These include loss of initial syllables or their replacement by a vowel; transfer of the feature 'roundness' from the vowel to an adjacent consonant (from which it might spread or migrate) – this resulted in the earlier three-vowel inventory being reduced to two, with later expansion, as noted previously; prestopping of certain nasals; and loss (or, as Koch (1997) has it, neutralization) of final vowels. Orthographically, in some dialects, all words are written with final *e*, representing schwa, whereas in others, final (predictable, often optional) vowels are not written, except, as *a*, in short words, in which they may be the stressed or even the only vowel. Thus, for example, earlier **nyina* 'sit' has become *n-* or *an-*; **ngali* 'we (dual, inclusive)' has become *il-*, *ayl-*, or

Table 1 Central Arrernte consonants

| Type | Peripheral | | Laminal | | Apical | |
|------------------|------------|-------|---------|----------|----------|--------------|
| | Bilabial | Velar | Dental | Alveolar | Alveolar | Postalveolar |
| Stop | p | k | th | ty | t | rt |
| Nasal | m | ng | nh | ny | n | rn |
| Prestopped nasal | pm | kng | thn | tny | tn | rtn |
| Lateral | | | lh | ly | l | rl |
| Tap | | | | | rr | |
| Glide | w | h | | y | | r |

aly-; *wama ‘snake’ has become *apme* or *apmwe*; and *munga ‘night’ has become *ingwe*. Breen and Pensalfini (1999) have argued that, contrary to the supposed universal situation that all languages have consonant–vowel (CV) syllables and that VC syllables can occur only in a language that also has CV, CVC, and V syllables, the sole underlying syllable shape in Arrernte is VC(C). Words that are consonant-initial on the surface have an underlying initial schwa. See Breen (2001) for a brief overview of the phonologies of the different dialects, and Koch (1997) for his view of the sound changes that have occurred.

In phonotactics, Arrernte is atypical in Australia in that it allows monosyllabic words and (surface) word-initial consonant clusters (homorganic or heterorganic). In most dialects, the vast majority of words are vowel-initial, mostly *a*-initial (the remainder having the underlying initial schwa, which never appears utterance-initially). The definition of the concept ‘word’ in Arrernte is problematic; units that are clearly words, or even phrases, can be incorporated into words, dividing them into parts that are clearly less than words (see Henderson, 2003).

Grammatically, Arrernte is typical of languages of most of Australia in the following ways:

- Nouns operate in an absolutive/ergative paradigm but pronouns are nominative/accusative, except that first- and second-person singular in eastern and northern dialects distinguish intransitive subject, transitive subject, and object.
- Pronouns have three numbers – singular, dual, and plural – and in some dialects distinguish exclusive from inclusive in first-person dual and plural (whereas others have lost this distinction but retain, with no function, the old exclusive marker).
- Cases are marked by suffixation.
- There is no grammatical gender.
- The rich verbal morphology includes a variety of compound types; verb suffixation marks tense, mood, aspect, associated motion, and, optionally, number of subject.
- Reduplication, of various types, is prominent in the grammar of the major lexical categories.
- Preferred constituent order is subject-object-verb (SOV), but this is frequently varied by pragmatic factors.

There is a complex interaction between kinship and grammar, although much of this is being lost. Society was, in the recent past, organized into four sections (called ‘skins’ in Aboriginal English) based on a division into two patrilineal moieties superimposed on a division of alternating generations. Not long before European settlement, a further division to form eight subsections diffused from groups in the northwest,

but did not reach to the easternmost or southernmost parts of the Arrernte area. Nonsingular pronouns can be marked according to the relationship of the persons concerned; thus, in Alyawarr, we have *ayla* ‘we (dual, inclusive, same section),’ *aylern* ‘we (dual, exclusive, same section),’ *aylak* ‘we (dual, inclusive, same moiety but differing by an odd number of generations, as father and child),’ *aylernak* (as *aylak*, but exclusive), *aylanth* ‘we (dual, inclusive, different moiety, as mother and child or husband and wife),’ and *aylernanth* (as *aylanth*, but exclusive). Kinship terms can be suffixed with morphemes derived from dative pronouns to indicate possessor; so, from *arreng* ‘father’s father and reciprocal’ we can have *arrengaty* ‘my father’s father,’ ‘my son’s child (I being male),’ ‘my brother’s son’s child,’ and so on; *arrengangkwa* ‘your father’s father,’ etc.; *arrengikwa* ‘his or her father’s father,’ etc.; *arrengalyew* ‘our father’s father, we being siblings,’ etc.; and *arrengalyewak* ‘father’s father of one of us (we being in the same moiety but differing by an odd number of generations, as father and child),’ etc. This last term, *arrengalyewak*, can be used by some speakers in the singular (‘father’s father of one or other of us’), but others could use it only if, say, there were two people who were ‘your and my father’s fathers.’ Each of the 27 possible suffixes can be used in this way. The following sentence is a relatively simple example in Antekerrenpenh, translated by the speaker (SS means ‘same section’; VOC, vocalic; ERG, ergative; 1DU, first-person dual; DAT, dative):

- (1) Angkwer-ey antyeny ayn-el-ayl-ek
 elder.sister-VOC old.man father-ERG-1DU.SS-DAT
 aherr atw-ern.
 kangaroo kill-PAST
 ‘Well sister, old dad’s killed a kangaroo.’

Another example, in Alyawarr, is from a children’s story (Summer Institute of Linguistics, 1996); the stories were the result of a linguist showing the (adult) language workers a series of drawings and asking them to make a story about the drawings. The first story in the book, about three boys who got lost, had the following sentence (DM means ‘different moiety’; 3PL, third-person plural):

- (2) Am-ayn-ew-anth-err-then
 mother-3PL-DAT-DM-PL-also
 ayn-ayn-ew-anth-err-then
 father-3PL-DAT-DM-PL-also
 nthw-ew-anem ampa ikwer-rnem
 look.for-PAST-then child 3SG.DAT-PL

This is translated as ‘Their mothers and fathers looked for the children’; the boys could have been two brothers and their cross-cousin. The same kinship

terms were used with ergative marking on the following page of the story. Note that, there is no number marking, but the use of the complex kinship terms seems perfectly natural and efficient (these and other complexities of kinship grammar are as yet unpublished, but see Breen (1998) and Green (1998)).

Bibliography

- Anderson V B (2000). Giving weight to phonetic principles: the case of place of articulation in Western Arrernte. Ph.D. diss., Los Angeles: University of California.
- Breen G (1998). The grammar of kinship in Central Australia. Manuscript, course given at the Australian Linguistic Institute, University of Queensland.
- Breen G (2000). *Introductory dictionary of Western Arrernte* [with cultural notes by J Pfizner]. Alice Springs: IA Press.
- Breen G (2001). 'The wonders of Arandic phonology.' In Simpson J, Nash D, Laughren M, Austin P & Alpher B (eds.) *Forty years on: Ken Hale and Australian languages*. Canberra: Pacific Linguistics. 45–69.
- Breen G & Pensalfini R (1999). 'Arrernte: a language with no syllable onsets.' *Linguistic Inquiry* 30(1), 1–25.
- Green J (1992). *Alyawarr to English dictionary*. Alice Springs: IAD Press.
- Green J (1998). Kin and country: aspects of the use of kinterms in Arandic languages. M.A. thesis, University of Melbourne.
- Hale K L (1962). 'Internal relationships in Arandic of Central Australia.' In Capell A (ed.) *Some linguistic types in Australia*. Sydney: University of Sydney. 171–185.
- Henderson J (1998). Topics in Eastern and Central Arrernte grammar. Ph.D. thesis, University of Western Australia.
- Henderson J (2003). 'The word in Eastern/Central Arrernte.' In Dixon R M W & Aikhenvald A Y (eds.) *Word: a cross-linguistic typology*. Cambridge: Cambridge University Press. 100–124.
- Henderson J & Dobson V (1994). *Eastern and Central Arrernte to English dictionary*. Alice Springs: IAD Press.
- Koch H J (1997). 'Pama–Nyungan reflexes in the Arandic languages.' In Tryon D & Walsh M (eds.) *Boundary rider: essays in honour of Geoffrey O'Grady*. Canberra: Pacific Linguistics. 271–302.
- Koch H J (2004). 'The Arandic subgrouping of Australian languages.' In Bowern C & Koch H (eds.) *Australian languages: classification and the comparative method*. Amsterdam/Philadelphia: John Benjamins. 127–150.
- Strehlow C (1907–1920). *Die Aranda-und Loritja-Stämme in Zentral-Australien. Veröffentlichungen aus dem Städtischen Völker-Museum Frankfurt am Main*. Frankfurt am Main: Joseph Baer.
- Strehlow T G H (1971). *Songs of Central Australia*. Sydney: Angus and Robertson.
- Summer Institute of Linguistics (1996). *Ampernem apererterrew*. Darwin, Australia: Summer Institute of Linguistics.
- Wilkins D P (1989). *Mparntwe Arrernte (Aranda): studies in the structure and semantics of grammar*. Ph.D. thesis. Canberra: Australian National University.

Artificial Languages

P O Bartlett, Vienna, VA, USA

© 2006 Elsevier Ltd. All rights reserved.

An artificial language is a language that has been deliberately designed for a purpose by one person or a small group of people over a relatively short period of time. (Adapted with permission from a definition by Richard K. Harrison, personal communication, 2004.) This definition, while serviceable, does lead to uncertain cases, such as whether pidgins should be considered artificial, being developed by small groups for a purpose, but usually pidgins are not considered to be artificial languages and will not be treated as such within the scope of this article. There is additionally the question of whether reduced languages such as Basic English (1930) are artificial. Also, this article pertains only to languages for interhuman communication and therefore does not address such constructs as computer programming languages.

Constructors can operate from any of several motives for designing a language. Some language designers intend that eventually their languages will replace an entire family of languages, such as Tutonish (1902) for the Germanic languages, or Ro (1906) for the entire world, considering that their languages would confer some overwhelming advantage to warrant replacing other existing languages.

Perhaps the most common design goal of artificial languages is international auxiliary languages, languages intended for use among people who do not have (or do not choose to use) any other language in common. Auxiliary languages, of which the best known but by no means the only is Esperanto (1887), may be intended to serve among localized areas (e.g., Guosa in Nigeria, 1965) or for the whole world. (Some have questioned whether replacement and auxiliary languages should be considered real languages. The experience of Esperanto, among others, tends to show that at least some such languages are

adequate for any level of discourse for which their users want to employ them. Also, there are individuals who have learned Esperanto from infancy in Esperanto-speaking homes. Therefore at least some auxiliary languages are real languages.)

There are authors who have designed languages, at highly varying levels of specification and completeness, for artistic use or to be part of a fictional or mythic world. Examples are the Elvish languages of J. R. R. Tolkien's Middle Earth and Klingon on the *Star Trek* series.

A few languages have been designed to test some or other linguistic hypothesis. The original motivation of James Cooke Brown's Loglan (1960) seems to have been to test the Sapir-Whorf Hypothesis.

From time to time, smaller or larger groups have constructed languages in order to communicate among themselves without their communications being readily intelligible to outsiders. Often such concealment languages, such as the Pig Latin of childhood, are modifications of existing languages.

Special communication needs are a sixth motivation for constructing languages. Some, such as a later adaptation of Blissymbolics (originally Semantography, 1949), are designed for communication needs of persons with physical and/or mental disabilities. Languages allegedly usable in psychoanalysis and psychotherapy, such as aUI (1962), are another venue. Additionally, intended communication, at least on a rudimentary level, with hypothesized extraterrestrial beings can give rise to a language.

Some individuals (and occasionally small groups) construct languages merely for enjoyment, as hobbies, just as some people construct model ships.

Finally, there may be miscellaneous occasions, such as altered religious and/or mental states, although one might question whether some such languages are constructed for a conscious purpose.

Artificial languages, and auxiliary languages especially, have various provenances. The Indo-European (IE) matrix of language designers seems to be the most common provenance of languages readily documented, that is, the designers themselves tend to be speakers of IE languages, and the products are heavily influenced by an IE substrate. In many instances, the languages have an intended primary audience of speakers of European languages, including speakers of non-IE European languages. However, some languages such as Afrihili (1970) have target audiences other than Indo-European speakers (although the languages themselves are often presented and described using an IE language). On the other hand, some auxiliary languages may have an IE base but have an intended audience worldwide.

The history of artificial languages, even in the West, is extensive, and only the briefest outline is possible, inasmuch as the number of auxiliary languages alone is in the hundreds spanning several centuries.

One of the earliest constructed languages in the West of which there is a record is the Lingua Ignota of St. Hildegard of Bingen (12th century). It comprised a 23-letter alphabet and about 1000 words. It is not entirely clear whether she intended it as an amusement, as an auxiliary language, or to express certain religious assertions, such as mystical states.

In the 13th–14th centuries, Ramón Llull wrote his *Ars Magna*, which he conceived as a perfect and universal language, especially for the religious conversion of non-Christians.

It was during the 17th century (and later) that the so-called *a priori* philosophical languages came to the fore, especially with the Real Character (1668) of Bishop John Wilkins in Great Britain. Perhaps the most notable characteristic of the philosophical languages is their basis in a classificatory scheme of (supposedly) all knowledge. Knowledge is broken into categories, and the vocabulary follows in almost mathematically combinatorial form from the classification.

During the 18th and 19th centuries, a number of artificial language proposals surfaced, such as the rather eccentric Solresol (1827) by (Jean) François Sudre, based on a seven-note musical scale, which did, nonetheless, gather some interest.

Volapük, invented in 1879 by the Catholic priest Johann Martin Schleyer, was the first artificial language designed as an auxiliary language to gain any substantial following. It was an *a posteriori* language, i.e., one in which the grammar and (especially) vocabulary derive from one or more existing languages, although word forms of Volapük were greatly modified from readily recognizable forms. The language enjoyed considerable initial enthusiasm throughout Europe and North America, although that enthusiasm quickly waned due to what some considered to be shortcomings in the language itself, factional infighting within the movement, and the rise of Esperanto. (There were, however, some derivatives of Volapük itself, and the language, in a revised form, did have some slight revival in the 20th century.)

Esperanto (1887), the brainchild of Ludwig Lazarus Zamenhof (spellings vary), has become the most successful, in relative terms, of all the artificial auxiliary languages to date. It has a largely Indo-European grammar with a rather agglutinative word-formation system. Estimates of the number of

Esperanto speakers differ widely and are controversial, ranging from a few tens of thousands to several million. Over the decades, people have raised various objections to Esperanto's structure, vocabulary, or orthography (which includes some accented letters unique to itself). Consequently, Esperanto has given rise to numerous derivatives, of which the only one to have any significant number of users at all has been Ido (1907).

Several artificial languages have the design goal of being naturalistic in terms of recognizability to speakers of west European languages. Notable among them have been Latino sine Flexione (1903) by Giuseppe Peano, a kind of Latin with most of the inflections stripped out, Occidental (1922) of Edgar de Wahl, and Interlingua (1951) of the International Auxiliary Language Association, Inc.

A few artificial languages have been known as logical languages, being based on predicate logic rather than on more common grammatical principles. Among these are Loglan (1960) and Lojban (1988) by the Logical Language Group, Inc.

Finally, there have been numerous artificial languages, too many and too varied to try to describe here even cursorily, that might be subsumed under the catch-all heading of just about anything under the sun. They have characteristics similar to those of languages all over the world.

Artificial languages have various features in both grammar and vocabulary, although the grammars of auxiliary languages (at least those developed by Westerners) often (although not always) tend to follow an Indo-European model.

A priori languages, first mentioned above, have two overlapping types. There are those, such as Wilkins's Real Character, Foster's Ro, or Elam's Oz (1932), which follow a classificatory system for vocabulary, as noted above. Such schemata are open to several criticisms:

- The totality of knowledge does not always fit neatly into a simple and single taxonomic schema.
- The taxonomic schema is dependent on the state of knowledge at the time of the creation of the schema.
- It can be difficult to fit new discoveries, taxa, and techniques into the schema, as the schema tends to be relatively closed.
- In practice there is a prodigious demand on the memory (and on the oral-aural channel) to retain the schema and to make fine distinctions (both semantic and oral).

Another use, however, of the term '*a priori*' is simply a reference to artificial languages whose

vocabularies are made up *ad hoc* and not derived from the vocabularies of existing languages. Some languages of this type (many examples could be cited) may have some internal structure to the vocabulary, primarily for mnemonic value, but do not follow a classificatory scheme as such.

A posteriori languages have their grammar and vocabulary bases in existing languages. The degree to which the vocabulary items are deformed varies widely.

There are also logical languages, as mentioned above. Their vocabulary may be *a priori* or at least partially *a posteriori*.

Auxiliary languages in particular can have different intended audiences and purposes. Some designers target their products largely for informal, personal use, such as among travelers and correspondents. On a wider scale, commercial and professional applications may come into purview. IALA Interlingua has seen some professional use in the past, but few languages seem to have yet found much widespread use in the commercial realm. Intergovernmental use, such as diplomacy and treaties, may be encompassed within the design of a language, although none have yet made significant inroads into this area. Different members of target audiences may have different assessments of the ease with which adult learners can acquire and use an artificial language.

Artificial languages in general (and not just auxiliary languages) differ markedly in their division of semantic space. Some have a rich vocabulary, making fine semantic distinctions, and others have a much more restricted vocabulary, depending on periphrasis to convey distinctions. Languages differ widely in how they handle (or even allow) unassimilated or partly assimilated foreign terms.

The issue of idiom often tends not to be treated extensively in the construction of auxiliary (and other artificial) languages. As a result, many users often import native idioms, impeding ready communication, or make conscious efforts to avoid idiom entirely. Of course, there is nothing to prevent a body of users from developing over time idioms unique to the user base itself.

Just why an auxiliary language does or does not have much use (in terms of speaker base) may depend on several factors. Not all of these factors are linguistic characteristics in and of themselves. Among them are:

- Propitiousness of circumstances, or 'right place at the right time.' In some language milieux, there is simply little felt need for an auxiliary language.

- Perception by prospective learners and users that the language itself is adequate for the task and sufficiently easily acquired by adult learners. This factor, although highly subjective, is operative. This factor can be called 'good enough.'
- A proposed international auxiliary language must have a stable enough base so that it is not always moving under the feet, so to speak, of would-be users. (Some language designers continue indefinitely to make changes.)
- A language proposal must be sufficiently dispersed to the attention of prospective users, with didactic material available.
- Proponents must have sufficient enthusiasm to work against social inertia.
- Proponents must have at least a minimally sufficient organization at some time to assist propagation.
- External events, such as wars or government favorable (or disfavorable) attention, may work for or against the spread of an auxiliary language.

Although much material exists for individual artificial languages, there are few comprehensive studies

of artificial languages in general. Most available material relates to international auxiliary languages, and some of that is on a popular level. Some of the works cited in the Bibliography contain further references for the interested reader.

Bibliography

- Eco U (1995). *The search for the perfect language*. Oxford: Blackwell Publishers.
- Harrison R K (1992). *Bibliography of planned languages (excluding Esperanto)*. Orlando, FL: R. K. Harrison.
- Large A (1985). *The artificial language movement*. Oxford: Basil Blackwell.
- Pei M (1958). *One language for the world*. New York: Devin-Adair.
- Schubert K (ed.) (1989). *Interlinguistics*. Berlin/New York: Mouton de Gruyter.
- Tonkin H & Johnson-Weiner K (eds.) (1986). *The idea of a universal language*. New York: Center for Research and Documentation on World Language Problems.

Assamese

P N Dutta Baruah, Central Institute of Indian Languages, Mysore, India

© 1994 Elsevier Ltd. All rights reserved.

Assamese is the principal vernacular and official language of Assam, a northeastern state of India, and is spoken by 10 million persons there and by 10 million more in Bangladesh. An Anglicized derivation of *ᱡᱷᱟᱨ* 'Assam,' Assamese refers to both the language and the speakers. Natives call it *ᱡᱟᱢᱥᱤᱞ* < *ᱡᱷᱟᱨ* + *iya* meaning 'belonging to.' A descendant of the Magadhan group of the Indo-Aryan family of languages, it shows affinity with modern Hindi, Bengali, and Oriya. Its formative period begins from the tenth century and written records in verse date from only the late thirteenth century, *prahlada charita* by Hem Saraswati being the earliest one. Developed from Brahmi through Devanagari, its script is similar to that of Bengali except the symbols for /r/ and /w/; there is no one-to-one phoneme-grapheme correspondence.

Its characteristic phonemic features include a voiceless velar fricative /x/, the alveolar fricatives /s/ and /z/, alveolar plosives, the alveolar nasal /ɳ/, only one /r/, and the intervocalic occurrence of /ŋ/. Characteristic

morphological features are: (a) gender and number are not grammatically marked; (b) there is lexical distinction of gender in the third person pronoun; (c) transitive verbs are distinguished from intransitive; (d) the agentive case is overtly marked as distinct from the accusative; (e) kinship nouns are inflected for personal pronominal possession, e.g., *deuta* 'father,' *deuta-r* 'your father,' *deuta-k* 'his father'; (f) adverbs can be derived from verb roots, e.g., *mon pokhila uradi ure* 'The mind flies as a butterfly flies'; (g) a passive construction may be employed idiomatically, e.g., *eko nuxuni* 'Nothing is audible.'

Syntactically it is non-distinct from its genetic relatives. Assamese has no caste dialects but a geographical dialect *kamrupi* with further sub-dialects. Written Assamese is almost identical with standard colloquial. An Assamese-based pidgin, Naga Pidgin or Nagamese, is spoken in Nagaland. Mutual convergence with neighboring Tibeto-Burman languages and Bengali spoken in Assam is noticeable in phonology and vocabulary. Its indigenous vocabulary is gradually falling into disuse in favor of Sanskritized forms. It stands unique among its genetic relatives in having developed historical and biographical prose as far back as the sixteenth century.

Bibliography

Grierson G A (1903). *Linguistic Survey of India*, vol. V, Part I. Calcutta: Office of Superintendent of Government Printing.

Kakati B K (1962). *Assamese – Its Formation and Development*. Gauhati: Lawyers Book Stall.

Pattanayak D P (1969). Oriya and Assamese. In Sebeok T A (ed.) *Current Trends in Linguistics*, vol. 5. The Hague: Mouton.

Census of India (1991) series. New Delhi: Office of the Registrar General.

Australia: Language Situation

P K Austin, School of Oriental and African Studies, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The languages spoken in Australia can be classified into the following:

- indigenous languages spoken by Aboriginal and Torres Strait Islander people;
- pidgins and creoles arising from language contact, primarily spoken by Aboriginal and Torres Strait Islander people and the descendants on Pacific Islander groups;
- community languages, including Australian Sign Language (Auslan) and the languages spoken by immigrant community groups and their descendants;
- Aboriginal English, primarily spoken by Aboriginal and Torres Strait Islander people;
- Australian English, the official language of the country and spoken as a first language by 90% of the population, with regional and social variation.

Aboriginal and Torres Strait Islander Languages

When Australia was colonized by Europeans in the late 18th century, it was home to approximately 250 indigenous Aboriginal and Torres Strait Islander languages (Dixon, 1980; Walsh, 1997; Angelo *et al.*, 1994; Austin, 1996), many of which are now either extinct, moribund, or endangered. Today, only 12 indigenous languages continue to be learned by children (McConvell and Thieberger, 2004), meaning that 95% of Australia's indigenous heritage has disappeared or is highly threatened. Recently there have been moves toward revitalization of Aboriginal languages (see below).

The languages spoken in the Torres Strait Islands fall into two groups: Meryam Mer, spoken in the

eastern islands, is related to Papuan languages to the north, and Kala Lagaw Ya, spoken in the western islands, is related to languages of the Australian mainland. For Tasmania, the existing sources are poor and it is difficult to say much definitively about the traditional indigenous language situation (Crowley and Dixon, 1981); however, much work has been done on reconstructing old sources (Crowley, 1993) and the Tasmanian Aboriginal Centre is promoting the revived language.

There has been growing awareness of Aboriginal and Torres Strait Islander languages among the general Australian population, and Aboriginal language courses are now taught in secondary schools in Victoria, South Australia (Nathan, 1996), and soon to be introduced in New South Wales. Bilingual education is also available in the Northern Territory, Queensland, and Western Australia, although programs are often threatened with funding cuts and lack of staff. Over the past 20 years, a number of Aboriginal-run Language Centres have been established throughout the country to collect language and culture information, prepare practical materials such as dictionaries and text collections, and to support local education and cultural revival initiatives. These grassroots organizations have been successful in mobilizing scarce resources in support of the languages. National bodies such as the Federation of Aboriginal and Torres Strait Islander Languages (FATSIL) have been set up, and Aboriginal languages have an increasing presence on the internet (see David Nathan's Aboriginal Languages Virtual Library website for sources). The Central Australian Aboriginal Media Association is also involved in broadcasting and recording and distribution of Aboriginal music. Since the 1980s, Aboriginal rock music bands, some of whom, such as Yothu Yindi, sing in indigenous languages, have become popular across Australia and internationally.

Although threatened by dominant Australian English, there are signs of indigenous language and cultural revival in South Australia (Amery, 2001)

and elsewhere. In 2003, the New South Wales government committed significant funds to supporting indigenous languages in that state and introducing them into the school system in the Languages Other than English (LOTE) program.

Language Relationships

The indigenous languages spoken across the southern two-thirds of the Australian continent plus eastern Arnhem land belong to a single language family called Pama-Nyungan, originally proposed by Kenneth Hale and Geoffrey O'Grady in the 1960s. Much descriptive and comparative work, especially in the last 10 years, has provided support for this family group (see Bowern and Koch, 2003 for the most recent sources, especially the extensive cognate materials given by Alpher in that volume). In the 'Top End' (the Kimberley, Daly River, and western Arnhemland), there is much more linguistic diversity, with some 20 language families having been identified (although recent research has increasingly argued that higher level groupings may also exist; see Evans, 2003). Whether all the languages are ultimately related as a single genetic family remains to be determined.

Linguistic Characteristics

Traditionally, Aboriginal groups were multilingual, as a result of exogamous marriage patterns, and individuals spoke several languages, while claiming primary allegiance to the tongue of their descent group. Languages also showed sociolinguistic variation: geographically different dialects, and special speech styles reflecting kinship and ritual relationships (see Walsh and Yallop, 1993).

Phonologically, languages generally lack fricatives and affricates, and there are contrastive stops at up to five points of articulation, with a nasal for each stop position, one or more laterals, a flap, a semiretroflex continuant, and two glides (see **Gamilaraay** and **Jiwarli** for further details). Stops and nasals contrast laminal and apical manners of articulation. There is usually no voicing contrast for stops (i.e., no contrast between *p* and *b*, for example). Most languages have just three vowels: high front *i*, high back *u*, and low *a*, with a phonemic length contrast found in about half the languages (Dixon, 1980). Some Cape York Peninsula languages have undergone historical sound changes introducing fricatives, prenasalized stops and additional vowel contrasts; Arandic languages of Central Australia are argued to have only two vowels and a contrast between rounded and unrounded consonants (see Breen in Simpson *et al.*, 2001).

The general phonotactic structure of word roots is CV(C)CV(C). Every word must begin with a single consonant and end in a vowel, or a restricted number of consonants. Some languages only allow vowel-final words (see **Jiwarli**). Word initially, in general only nonapical stops and nasals, and the two glides are found. Word medially, there are limited consonant clusters, primarily homorganic nasal plus stop, and apical nasal or lateral plus peripheral stop (*p* and *k*). Vowel clusters are not found, though Vowel-Glide-Vowel sequences are possible. Word stress is generally not phonemic and predictable from the phonological shape of words (see **Gamilaraay** for examples).

Languages of the Pama-Nyungan (PN) group are entirely suffixing in their morphology; non-Pama-Nyungan (non-PN) languages may show both suffixes and prefixes, and tend to be head-marking rather than dependent-marking. There are two major word classes: nominals and verbs, with nominals in PN languages typically showing rich systems of case-marking (in non-PN case-marking is often absent) and verbs marking tense/aspect/mood and dependent clause categories. Nominals can be subdivided into substantives (that cover both noun and adjective concepts in a language like English), pronouns, and demonstratives. Minor word classes include adverbs, particles, and interjections.

Nominals in PN languages typically inflect for case, with the syntactic functions of intransitive subject (S), transitive subject (A), and transitive object (P) showing a split-ergative pattern of syncretism in the case forms determined by animacy:

- for pronouns S and A fall together as a single (unmarked) form with P different (making nominative-accusative case marking);
- for other nominals, S and P fall together as a single (unmarked) form with A different, making ergative-absolutive case marking.

In some languages, some nominal categories (e.g., animate nouns) show a three-way contrast distinguishing S-A-P. In non-PN languages, there are typically systems of verb affixation encoding agreement with verb arguments; this agreement may also reflect gender categories of the nominals.

The following cases are also typically found in PN languages:

- dative, marking alienable possession and direction toward a place;
- locative, coding location in a place;
- ablative, coding direction from a place, and cause.

Australian languages typically have complex systems of nominal word-building morphology that involves suffixation between the root and case inflection. Categories encoded in word-building morphology include number, having, and lacking. Some non-PN languages encode gender on nouns via affixation.

Pronouns generally distinguish three persons and singular, dual, and plural number; in the first person nonsingular, there is an inclusive–exclusive contrast in about half the languages. Some languages also show bound pronouns, often these are reduced forms of the free pronouns and in PN languages are suffixed to particular elements of the clause (Dixon, 1980).

Verbs morphologically distinguish between main verb and dependent verb inflections. Main verbs encode tense and mood categories, while dependent verbs occur in hypotactically linked clauses and mark relative tense (and in some central Australian languages also switch-reference; see below). There are typically a number of verb conjugations that are morphologically determined but may show some correlations with transitivity (Dixon, 1980). Verbs show productive word-building morphology, including affixes that indicate aspectual categories or change in transitivity (detransitivizing and transitivity processes). Generally passive forms are not found, though some eastern Australian languages have anti-passive derivations. Non-PN languages show agreement via affixation on the verb. The minor categories of adverb, particle, and interjection show no morphological variation. All languages also have affixes that attach to words of any category, typically encoding discourse status, evidentiality, and other pragmatically based meanings.

A common pattern in many Australian languages (see **Jiwarli**) is for word order to be relatively free and hence to find all possible orders of Subject, Object, and Verb, as well as separation of nouns and adjectives referring to a single entity (with case agreement indicating common reference). Similarly, possessors (in dative case) may precede or follow the alienable possessed noun. Free omission of nominals whose reference is clear from the context is also common. Australian languages have become famous for their ‘nonconfigurational syntax.’

Interclausal syntax shows a degree of variation; some languages (see **Gamilaraay**) place little restriction on linking of clauses, while others such as **Dyirbal** have ‘ergative syntax’ where the linked clauses must share coreferential absolutive (S or P) nominals. Many central Australian languages have switch-reference where cross-clausal identity or nonidentity of subjects (S or A) is encoded on the dependent verb. Non-PN languages tend to make use of parataxis in clause linkage.

Particles in Australian languages tend to have scope over the whole clause and encode such semantic concepts as polarity (affirmation versus negation) and mood (possibility, negative imperative, etc.).

Pidgins and Creoles

Australia has a number of English-based pidgins and creoles as a result of language contact between the indigenous languages and the English of the colonizers, beginning in the late 18th century. A range of geographically diverse forms have been and are found, including Sydney-pidgin (extinct since the 19th century; Troy, 1990), Kriol of the ‘Top End,’ Cape York Creole (Crowley and Rigsby, 1979), and Broken or Blaikman Tok of the Torres Strait islands (see Schnukal in Angelo *et al.*, 1994). Kriol is now the native language of some 30 000 speakers in northern Australia.

The various creoles show clear influence from Australian indigenous languages both lexically and structurally (e.g., distinguishing singular, dual, and plural pronouns, and inclusive–exclusive reference in the nonsingular). They also share many characteristics with Pacific pidgins and creoles such as Tok Pisin and Bislama.

The descendants of Pacific islanders removed to Australia in the 19th century to work on sugar plantations in Queensland spoke Pacific pidgins and creoles – these are now being replaced by Aboriginal English.

Community Languages

As a result of on-going immigration of non-English speakers into Australia, some 200 languages have been added to the linguistic ecology of the country (see Clyne, 1991; Clyne and Kipp, 1997). The distribution of these ‘community languages’ varies regionally, especially between the major urban centers, e.g., Melbourne adolescents show dominance of Italian and Greek (reflecting immigration after the Second World War), while Sydney shows dominance of Arabic and Chinese languages (reflecting more recent immigration from the middle East and South-east Asia). All community languages are undergoing shift to English (Clyne and Kipp, 1997), though to varying degrees in different communities (e.g., more highly among Dutch than Poles or Maltese and Turks). Community languages are widely taught in schools (as LOTE), and bilingual education (including immersion programs) is available in some languages. Local governments in Australia, particularly in the urban centers, pay attention to community languages and provide services and information in a

range of languages. There is a system of registration for interpreters and translators, and strong infrastructure of telephone and court interpreting services for non-English speakers.

An important community language is Australian Sign Language (Auslan), which is widely used in the deaf community, and differs in significant ways from American Sign Language (ASL) and British Sign Language (BSL). After being ignored for a long time, research and publications on Auslan have appeared over the past 15 years (see Johnston, 1989, for example) and an active program of documenting Auslan is underway. Because of early diagnosis of deafness and the widespread use of cochlear implants in deaf children, the number of native Auslan signers has shown a dramatic decline in recent years; the language is currently endangered.

Aboriginal English

Aboriginal English is a particular form of Australian English primarily spoken by Aboriginal and Torres Strait Islander people. It is spoken as a first or second language and is a continuum that ranges from varieties that resemble pidgin or creole English to those more like nonstandard Australian English (Eagleson, 1983; Eades, 1991; Kaldor and Malcolm, 1991). Aboriginal English in rural settings shows substrate influence in articulation (having apico-dental retroflex articulations) and replacement of fricatives with stops, for example), lack of copula, lack of number marking and *bin* as a past tense marker. In urban settings, Aboriginal English shows many features found in nonstandard varieties across the world, such as multiple negation, and nonstandard verb agreement; however, there are lexical and pragmatic features (Eades, 1991) that are distinctive. Even in regions such as Sydney and Melbourne where the indigenous languages ceased to be spoken in the 19th century, Aboriginal English contains lexical items derived from the indigenous languages such as *koorie* 'Aboriginal person' and *goom* 'alcohol.'

Australian English

A distinctive Australian variety of English (AustEng) is spoken by 90% of the 20 million inhabitants of the continent, with regional and social variation. AustEng has its origins in the English dialects brought by mainly English and Irish settlers in the 18th and 19th centuries, to which have been added the speech of immigrants from all over the world. Long regarded as a substandard form of speech and lacking prestige (Turner, 1994), AustEng has become accepted over

the past 20 years and has been codified in dictionaries (including the Macquarie Dictionary in various versions dating from 1981, also now with a strong web presence, and the Australian National Dictionary), is used in English language teaching in Australia, and has been popularized in textbooks (e.g., Burrage and Mulder, 1998). It is now the prestige variety of English-language broadcasting. Like most other varieties of English, AustEng is currently being subjected to influence from American English, especially in the lexicon, but also in pronunciation (Burrage and Mulder, 1998).

Australian English shows a large number of loan words from indigenous languages (the Australian National Dictionary records over 400), especially for the distinctive flora and fauna of the country, and for place names, e.g., *kangaroo*, *billabong*, *waratah*, and *galah*, or *Woomooloo* and *Mordialloc* (see Dixon *et al.*, 1990 for other examples). Other sources of distinctive lexical materials include English dialects, convict slang, and rhyming slang, e.g., *Joe Blake* for *snake*, as well as locally developed terms, e.g., *outback*.

AustEng shows a degree of regional variation, particularly in vocabulary and pronunciation. Lexical variation has been well researched and increasingly documented in the dictionaries, while variation in pronunciation has been less studied. Among features that show geographical differences are [æ] vs. [a] in *graph* or *dance*, postvocalic vocalization of *l* (in words like *eagle*), lowering of low front [e] (in words like *Mel*, *helicopter*) and bisyllabification of past participles (so that *grown* sounds like *grow-en*).

Social variation in Australian English has been well studied since Mitchell and Delbridge (1965) established the categories of Broad, General, and Cultivated Australian. The differences are particularly clear phonetically in vowel nuclei, especially the diphthongs of *face*, *price*, *goat*, and *mouth* (see Harrington *et al.*, 1997). Table 1 below (from Melchers and Shaw, 2003: 105, based on Wells, 1982) shows the variants of Australian English vowels in comparison to Received Pronunciation.

Melchers and Shaw (2003: 104) list the following as especially salient features of AustEng:

- front [a:] in *palm*, and *start* (shared with New Zealand English)
- wide diphthongs in *fleece*, *face*, *price*, *goose*, *goat*, and *mouth*;
- close front vowels, in *dress*;
- extremely productive use of two noun suffixes *-ie* and *-o*,

Table 1 Australian English vowels

| Broad | General | Cultivated | Key word | RP |
|---------|-------------------|------------|----------|----|
| | ɪ | | kit | ɪ |
| | e | | dress | e |
| | æ | | trap | æ |
| | ɐ | | lot | ɐ |
| | ʌ | | strut | ʌ |
| | ʊ | | foot | ʊ |
| | aː | | bath | aː |
| | ɒ | | cloth | ɒ |
| | ɜː | | nurse | ɜː |
| əʊ | + i | ii | fleece | iː |
| ʌʔ _ aʔ | ʌʔ | ɛʔ | face | eʔ |
| | aː | | palm | aː |
| | ɔː | | thought | ɔː |
| ʌːʊ-aːʊ | ʌʊ | öʊ | goat | əʊ |
| əːʊ | ʊʊ | ʊʊ | goose | uː |
| ɒːʔ | ɒʔ | aʔ | price | aʔ |
| | ɔʔ | | choice | ɔʔ |
| ɛːʊ | æʊ | aʊ | mouth | aʊ |
| | ɪə - ɪə - iː | | near | ɪə |
| | eə | | square | ɛə |
| | aː | | start | ɒː |
| | ɔː | | north | ɔː |
| | ɔː | | force | ɔː |
| | ʊə - ɔː - ʊə - ʊː | | cure | ʊə |

- use of *she* as a generic pronoun, e.g., *she'll be right* 'it's fine';
- highly characteristic vocabulary, some drawn from indigenous languages, some from British dialect slang, and other elements locally developed.

Note also that AustEng differs from RP in having schwa in unstressed syllables, intervocalic voicing and flapping of *t*, and shares with it lack of post-vocalic *r* found in American and Canadian English. A distinctive high rising terminal intonation contour, noticed by Mitchell and Delbridge (1965) and investigated in depth for Sydney speech by Horvath (1985), is characteristic of female, teenage, and lower working class speech.

Morphologically, AustEng is characterized by a high degree of clipping, e.g., *uni* for *university*, *Oz* for *Australia*, which may or may not be combined with highly productive suffixation of *-ie* or *-o*, as in *Salvos* for *Salvation Army*, *maggie* for 'magpie,' *sunnies* for *sun glasses* and *lippie* for *lipstick*.

Bibliography

- Amery R (2001). *Warrabarna Kurna! Reclaiming an Australian Language*. Adelaide: Swets and Zeitlinger.
- Angelo D et al. (1994). *Australian phrasebook*. Melbourne: Lonely Planet.
- Austin P (1996). 'Australia.' In Comrie B, Matthews S & Polinsky M (eds.) *The atlas of languages: the origin*

- and development of languages throughout the world*. London: Quarto.
- Blake B J (1996). *Australian Aboriginal languages*. Brisbane: University of Queensland Press.
- Bowern C & Koch H (eds.) (2003). *Subgrouping and Australian languages*. Amsterdam: John Benjamins.
- Burridge K & Mulder J (1998). *English in Australia and New Zealand*. Melbourne: Oxford University Press.
- Clyne M (1991). *Community languages: the Australian experience*. Cambridge: Cambridge University Press.
- Clyne M & Kipp S (1997). 'Language maintenance and language shift: community languages in Australia, 1996.' *People and Place* 5(4), 19–27.
- Crowley T (1993). 'Tasmanian Aboriginal language: old and new identities.' In Walsh M & Yallop C (eds.) *Language and culture in Aboriginal Australia*. Canberra: Aboriginal Studies Press. 51–71.
- Crowley T & Dixon R M W (1981). 'Tasmanian.' In Dixon R M W & Blake B J (eds.) *Handbook of Australian languages*, vol. 2. Canberra: The Australian National University Press. 394–421.
- Crowley T & Rigsby B (1979). 'Cape York Creole.' In Timothy Shopen (ed.) *Languages and their status*. Cambridge: Winthrop Publishers. 153–207.
- Delbridge M & Delbridge A (1965). *The pronunciation of English in Australia*. Sydney: Angus and Robertson.
- Dixon R M W (1980). *The languages of Australia*. Cambridge: Cambridge University Press.
- Dixon R M W & Blake B J (1979–1983). *The Handbook of Australian languages*. Canberra: ANU Press.
- Dixon R M W, Ramson W S & Thomas M (1990). *Australian Aboriginal words in English: their origin and meaning*. Melbourne: Oxford University Press.
- Eades D (1991). 'Aboriginal English: an introduction.' *Vox: Journal of the Australian Advisory Council on Languages and Multicultural Education (AACLAME)* 5, 55–61.
- Eagleson R (1983). 'English in Australia and New Zealand.' In Bailey R W & Görlach M (eds.) *English as a world language*. Ann Arbor: University of Michigan Press. 415–438.
- Evans N (2003). *The northern languages of Australia*. Canberra: Pacific Linguistics.
- Harrington J, Cox F & Evans Z (1997). 'An acoustic study of broad, general and cultivated Australian English vowels.' *Australian Journal of Linguistics* 17(2), 155–184.
- Horvath B M (1985). *Variation in Australian English: the sociolects of Sydney*. Cambridge: Cambridge University Press.
- Johnston T (1989). *Auslan dictionary: a dictionary of the sign language of the Australian deaf community*. Sydney: Deafness Resources Australia, Ltd.
- Kaldor S & Malcolm I G (1991). 'Aboriginal English – an overview.' In Romaine S (ed.) *Language in Australia*. Cambridge: Cambridge University Press. 67–83.
- McConvell P & Thieberger N (2004). *State of indigenous languages in Australia 2001*. Australia: State of the Environment Second Technical Paper Series (No.1 Natural and Cultural Heritage).

- Melchers G & Shaw P (2003). *World Englishes*. London: Arnold.
- Nathan D (ed.) (1996). *Australia's indigenous languages*. Adelaide: Senior Secondary Assessment Board of South Australia [also includes CD-ROM].
- Schmidt A (1990). *The loss of Australia's Aboriginal language heritage*. Canberra: AIATSIS.
- Simpson J, Nash D, Laughren M, Austin P & Alpher B (eds.) (2001). *Forty years on: Ken Hale and Australian Aboriginal languages*. Canberra: Pacific Linguistics.
- Troy J (1990). *Australian Aboriginal contact with the English language in New South Wales 1788–1845*. Canberra: Pacific Linguistics.
- Turner G W (1994). 'English in Australia.' In Burchfield R (ed.) *The Cambridge history of the English Language V*. Cambridge: Cambridge University Press.
- Walsh M (1997). 'How many Australian languages were there?' In Tryon D & Walsh M (eds.) *Boundary rider: essays in honour of Geoffrey N. O'Grady*. Canberra: Pacific Linguistics C-136. 393–412.
- Walsh M & Yallop C (eds.) (1993). *Language and culture in Aboriginal Australia*. Canberra: Aboriginal Studies Press.
- Wells J C (1982). *Accents of English, vols I–III*. Cambridge: Cambridge University Press.

Relevant Websites

- <http://www.dnathan.com/vlibrary> – David Nathan's Aboriginal Languages Virtual Library website.
- <http://www.macquariedictionary.com.au> – Macquarie dictionary website.
- <http://www.fatsil.org> – FATSIL.

Australian Languages

B J Blake, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Archeological evidence indicates that Australia has been inhabited by humans for over 50 000 years. At the time of the establishment of the first British colony at Port Jackson (Sydney), in 1788, there were about 250 different languages spoken on the continent. Estimates of the Aboriginal population at that time vary from the low figure of 300 000 to several times that number. Over a period of a little more than 100 years, Europeans took over the whole country, killing a large proportion of the indigenous population in the process. Today only 60 or so Aboriginal languages are still spoken, and as few as 20 or so are likely to be spoken a generation from now.

For almost all the native languages, we have some record, though in some cases only a brief English–Aboriginal word list. Grammatical information is available for approximately 100 languages, the bulk of it having been collected since the 1960s, in many cases from the last speakers.

Classification

Capell classified Australian languages typologically into two groups: suffixing and prefixing, the latter group being confined to an almost continuous area in the north of the continent (see **Figure 1**). In the suffixing group, all affixes are suffixes, while in

the prefixing group there are some prefixes, mainly pronominal forms for subject and object (Capell, 1956: 31–60). The suffixing languages are predominantly agglutinative, but in the prefixing languages there is more fusion, mainly in the pronominal and other prefixes to the verb.

The languages of the mainland are generally thought to be related, since certain roots are widespread. These include lexical roots, such as *na* 'to see,' *mil* 'eye,' and *yan* 'to go,' and grammatical roots, such as *nga-* 'first person,' *nu* 'he,' and *ku* 'dative case marker'. In 1966 O'Grady, Wurm, and Hale produced a classification that recognized 29 'families' (O'Grady *et al.*, 1966a; O'Grady *et al.*, 1966b), but more recent work by various scholars has demonstrated that the figure could be reduced to as few as a dozen or so. The basis of the classification was lexicostatistical, and 'family' in this context meant a group of languages that could be linked on the basis of any member's sharing 15 percent or more of basic vocabulary with any other member.

A notable feature of the O'Grady *et al.* (1996a,b) classification is that one family, the Pama-Nyungan family, covers most of the mainland except for the Kimberleys and the Top End. It coincides roughly with the suffixing languages, taking in the Yolngu languages of northeast Arnhem Land, which represent an enclave of suffixing among the prefixing languages. The name Pama-Nyungan is derived from *pama* 'man' in the northeast of the continent and *nyunga* 'man' in the southwest.

Blake showed that between Pama-Nyungan and the other (Northern) languages, there are some

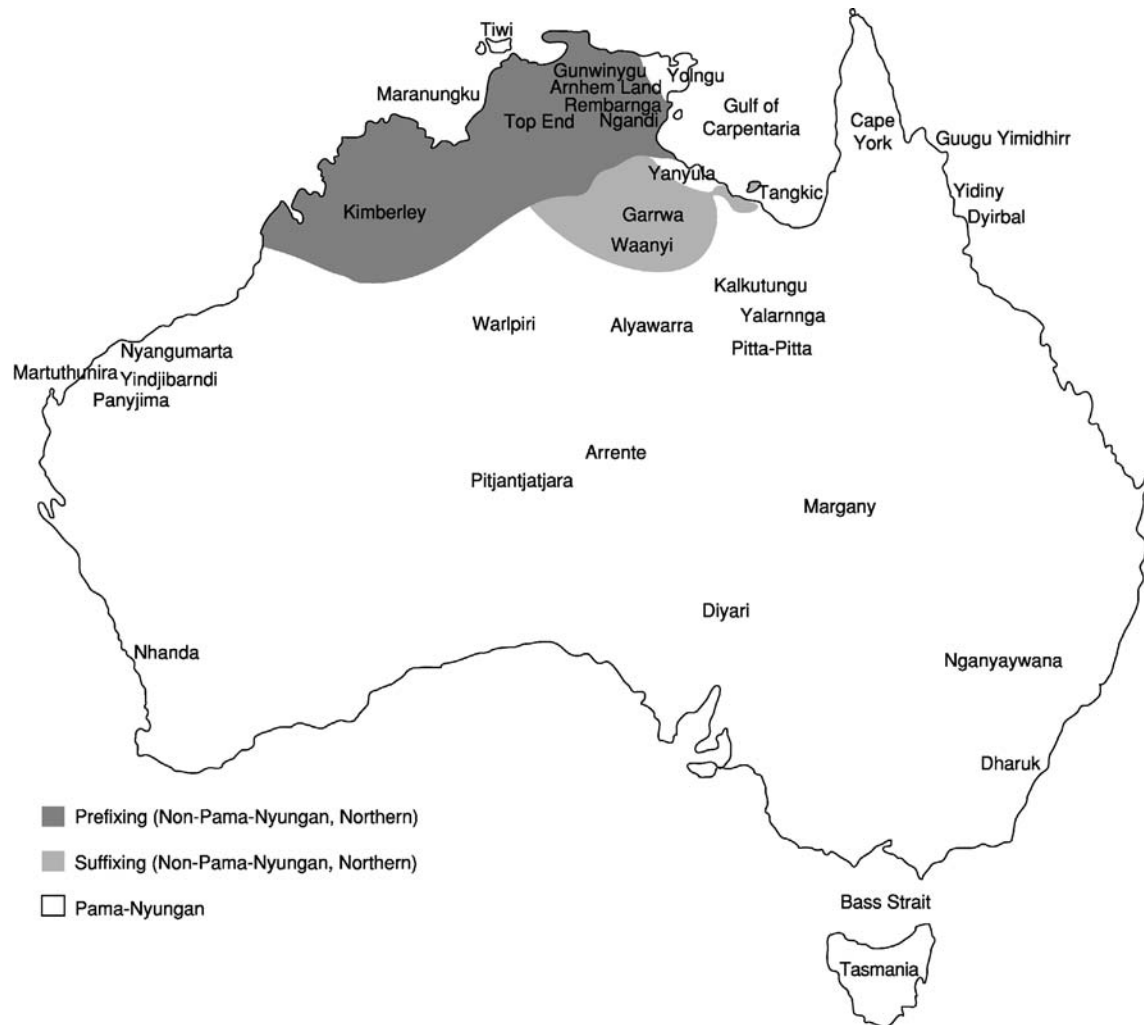


Figure 1 Pama-Nyungan and northern Australian languages.

consistent differences in the forms of some pronouns. For instance, while most Pama-Nyungan languages have a first person dual pronoun *ngali*, this is absent from the Northern languages, and while most Pama-Nyungan languages reflect a second singular **ngin*, a majority of Northern languages reflect **ngin'* with a palatal nasal as the third segment. A number of Pama-Nyungan languages have a third person pronoun root *nhu-*, whereas the Northern languages have *nu-* (Blake, 1988: 13). Blake's classification involved some reclassification, taking the Tangkic languages of the Gulf of Carpentaria to be Northern, and Yanyuwa to be Pama-Nyungan. Garrwa (Garawa) and Waanyi (Wan[j]i) are two languages with some Northern and some Pama-Nyungan pronouns.

Evans demonstrated that there is a regular correspondence between Pama-Nyungan and the Northern languages, reflecting a phonological change in Pama-Nyungan in which initial apicals (*t*, *n*, *l*) merged with

laminals (dental or palatal), the *nhu-/nu-* correspondence in the third person singular pronoun being part of the evidence for this change (Evans, 1988: 98–100).

While Blake and Evans provided evidence for a revised Pama-Nyungan that went beyond the lexicostatistical, in his recent book of Australian languages, Dixon argued strongly against the existence of Pama-Nyungan. He argued that the pronouns that characterize so-called Pama-Nyungan such as *ngali* 'we two' have diffused. He showed that the original lexicostatistical classification was flawed and that the shift in initial apicals to laminal did not coincide exactly with Pama-Nyungan. He also pointed out that no fauna or flora terms had been reconstructed that could be attributed to Proto-Pama-Nyungan (Dixon, 2002). Nevertheless, Australianists have so far not been convinced by Dixon's arguments (see, for instance, the papers in Bower and Koch, 2004).

As noted in this article, the languages of the mainland look as if they are related, though Dixon was pessimistic about the prospects of demonstrating this by the comparative method. There are several factors militating against reconstructing anything like Proto-Australian: the enormous time depth, demonstrable diffusion, and paucity of data, particularly for the southeast, which was taken over by Europeans early and was heavily settled.

It has not been possible to relate the languages of Tasmania to those of the mainland. Tasmania was cut off from the mainland about 14 000 years ago, when the earth warmed as it slowly emerged from the last Ice Age and the sea level rose, resulting in an unnavigable strait (Bass Strait) between Tasmania and the mainland. Given a time depth of 14 000 years for the period of separation, it is likely that any evidence of a genetic connection would have been obliterated.

It has likewise not been possible to establish a genetic connection between any Australian language, whether from the mainland or Tasmania, and any language from elsewhere.

Phonology

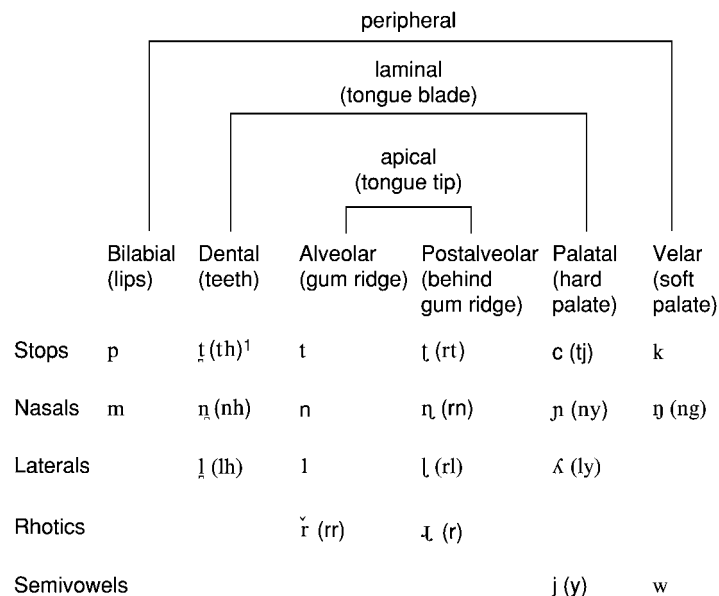
In Europe the phonologies of English, French, German, Italian, and Polish are quite different, but the mainland languages of Australia tend to be similar in their inventory of phonemes and in their phonotactics (word shapes). All Australian languages have stop sounds, but there is typically only one set, represented

either by *p*, *t*, *k*, etc., or by *b*, *d*, *g*, etc. Normally five or six stops are found: labial (*p*), apico-alveolar (*t*), apico-postalveolar or retroflex (represented here by *rt*), dorso-velar (*k*), and one or two laminal stops. Where there is one laminal stop, the pronunciation may range from dental to palatal, and by convention this stop is represented as palatal (*tj*). Where the dental and palatal stops are phonemically distinct, the dental is usually represented as *th*. Corresponding to each stop is a nasal. There is always one lateral (*l*), but there may also be dental (*lh*), palatal (*ly*), or retroflex (*rl*) laterals. Commonly there are two rhotics: a glide often described as retroflex and a flap, or trill. These are represented here by *r* and *rr*, respectively. All Australian languages have a labio-velar glide (*w*) and a palatal glide (*y*). **Figure 2** displays the consonants commonly found in Australian languages.

The majority of Australian languages have only three vowels (*i*, *a*, and *u*), though often there are long and short versions, which gives effectively six vowels. Some languages have *e* or *o* or both.

Words in Australian languages usually have more than one syllable, and more often than not they end in a vowel.

Although Australian languages right across the continent tend to have quite similar phonological systems, a few languages in a number of quite separate areas have undergone a series of phonological changes involving the loss of initial consonants or even whole syllables. In a number of Pama-Nyungan languages, there is a word *kumpu* for 'urine.'



1. Letter in parentheses are in common use and are used in this article.
Where voiced symbols appear in sources, they have been retained.

Figure 2 Consonants.

In Nhanda (western Australia), the initial consonant has been lost to yield *umpu*; in some languages of Cape York, the first syllable has been lost to give *mpu*; and in the Arandic languages of central Australia, the form *mpwa* occurs, the *k* having been lost and the *u* being reflected as labialization of the *mp* cluster. The effect of these changes has been to make some languages look quite atypical, and at one stage certain languages, such as Nganyaywana, were thought to be unrelated to other mainland languages, because cognate forms could not be readily recognized.

Morphology and Syntax

Inflection apart, words may be simple, compound, or reduplicated. In Pitta Pitta, for instance, *ngampa-manha* (stomach-bad) is ‘sad,’ and reduplicated forms of *ngapu* ‘water,’ *mayi* ‘dirt,’ and *maka* ‘fire’ yield *ngapu-ngapu* ‘wet,’ *mayi-mayi* ‘dirty’ and *maka-maka* ‘hot.’ The most common means of deriving new words is via suffixes. An almost ubiquitous feature of Australian languages is the presence of a suffix for ‘having’ and a suffix for ‘lacking,’ though the actual forms employed vary a good deal from language to language (see [1] and [3] for examples). In Pitta Pitta, for instance, we find forms like *kangamaru* (alcohol-having) ‘intoxicated’ and *nhupu-yaku* (spouse-lacking) ‘unmarried.’ Most languages have suffixes to mark the derivation of nouns from verbs and vice versa. In Diyari *wirlpa-nganka* ‘to make a hole’ is formed from the noun *wirlpa* ‘hole,’ and from this stem can be derived the noun *wirlpa-mganka-ni* ‘opener.’ Most languages have a suffix to mark the derivation of intransitive verbs from nouns, often with an inchoative sense. In Dieri we find formations such as *kilpa-rrri* ‘become cool’ and *yapa-rrri* ‘become afraid.’ Causatives of intransitive verbs are also common as in Diyari *pali-ma* ‘to extinguish a fire,’ from *pali* ‘to die.’ The majority of Australian languages express reflexive and reciprocal notions by using a derived intransitive verb. In Diyari we find *muduwa* ‘to scratch’ (transitive) and *muduwa-thadi* ‘to scratch oneself.’ Note the *d* in these words. Diyari has a voicing contrast in apical stops.

In the Pama-Nyungan languages, all derivational and inflectional affixes are suffixes. Nouns are marked for case, and verbs are marked for categories such as aspect, tense, and mood. In some languages, case concord extends from the head noun to its dependents; in others, it occurs only on the final word in the noun phrase.

With only a handful of exceptions, nouns in Pama-Nyungan languages take ergative case marking when functioning as the agent of a transitive verb (A) and zero case marking when functioning as the

sole argument of an intransitive predicate (S) and a direct object (O). The following examples are from Margany, a language of southwestern Queensland.

- (1) Nguda barndin-bayi.
dog dirt-having
‘The dog is dirty.’
- (2) Nguda-nggu yurdi gamba-nhi.
dog-ERG meat bury-PRES
‘The dog is burying the meat.’

On the other hand, in most Pama-Nyungan languages pronouns serving as S or A are treated alike (normally the bare stem is used, at least with nonsingular pronouns), while a pronoun in O function takes accusative case marking. This, too, can be illustrated from Margany.

- (3) Ngali bulu-idba.
we.two food-LACKING
‘We have no food.’
- (4) Gara ngali nhaa-nhi ina-nha.
not we.two see-PRES you-ACC
‘We can’t see you.’

Typically there is a dative case, an allative (‘to’), a locative (‘at’), an ablative (‘from’), frequently a genitive, and sometimes a causal or aversive that can cover cause, as in ‘I’m sick from (eating) bad meat,’ or what is to be avoided, as in ‘Keep away from the fire.’ The paradigms of Margany case forms displayed in Table 1 are typical with respect to both forms and categories. However, there is one idiosyncratic difference. The ergative case marker covers not only instrumental function, as it does in the majority of Pama-Nyungan languages, but also the causal or aversive sense alluded to in this article. In this function it can occur with pronouns and contrasts with the unmarked form used for the agent of a transitive verb.

A feature of case marking in Australian languages is the prevalence of double case marking. This is found, for instance, where a genitive-marked dependent of a noun displays case concord with its head, as in Margany.

Table 1 Margany case marking

| | | |
|--------------|-------------------|-----------------------|
| English | stone | we two |
| nominative | <i>barri</i> | <i>ngali</i> |
| ergative | <i>barringgu</i> | <i>ngali</i> |
| accusative | <i>barri</i> | <i>ngalinganha</i> |
| genitive | <i>barrigu</i> | <i>ngalingu</i> |
| dative | <i>barrigu</i> | <i>ngalingun.gu</i> |
| allative | <i>barridhadi</i> | <i>ngalingundhadi</i> |
| locative | <i>barringga</i> | <i>ngalingunda</i> |
| ablative | <i>barrimundu</i> | <i>ngalingunmundu</i> |
| instrumental | <i>barringgu</i> | <i>ngalingundu</i> |

- (5) Ngaya waban-gu ngali-ngu-ngga bama-ngga.
 I go-PURP we-GEN-LOC bro-LOC
 'I'm going with our brother.'

In about two-thirds of Australian languages, there are bound pronominal representations, either clitic pronouns or inflection, for subject (S and A) and object (O), and in a few languages there are forms for other complements or adjuncts, such as recipients or beneficiaries. In the suffixing languages, these pronominal elements are suffixed to the verb or to the first constituent in the clause.

Examples (6), (7), and (8), from Pitjantjatjara, illustrate the contrast between *-rna* the S/A(subject) form and *-rni* the O form. In this language, the bound pronouns are enclitic to the first constituent in the clause.

- (6) Munu-rna purta kapi-ku kutju a-nkuku?
and-I.SUBJ QUERY water-DAT alone go-FU
 'And should I go for water alone?'
 (7) Purnu-rna mantji-nu.
wood-I.SUBJ get-PT
 'I got the wood.'

Example (8) illustrates the double object construction, with a verb for 'give' in which the noun serving as patient object is unmarked, while the recipient is represented by both an accusative marked pronoun and a bound object pronoun that is enclitic to the first constituent. There is no overt form for third person subject.

- (8) Minyma-ngku-rni mayi ngayu-nya u-ngu.
woman-ERG-1SG.O bread 1SG-ACC give-PT
 'The woman gave me bread.'

In some languages, there is a detransitivized construction in which the agent of a two-place verb is encoded as S and the patient is expressed in the dative or some other oblique case. The following pair of sentences from Pitta Pitta (Queensland) illustrate the normal transitive construction and the derived intransitive construction, which, following Silverstein (1976), is generally known as the antipassive (AP). Pitta Pitta and some other related languages of western Queensland are unusual in that they have both ergative and accusative marking on all nouns and pronouns.

- (9) Pithi-ka nga-thu ina.
hit-PT I-ERG you.ACC
 'I hit you.'
 (10) Pithi-li-ya ngantja in-ku.
hit-AP-PRES I you-DAT
 'I feel like hitting you.'

The antipassive has a different semantic function in different languages, but it always signals some kind of

reduced semantic transitivity. In Pitta Pitta, it signals desiderative aspect.

Pitta Pitta uses a construction similar to the antipassive in the future tense. The verb is unmarked, there being neither the derivational antipassive nor the past or present inflection, and the subject (S or A) bears a special future subject inflection.

- (11) Pithi nganyu in-ku.
hit.FU I.FU.SUBJ you-DAT
 'I'll hit you.'

Pama-Nyungan languages are generally referred to as 'ergative'; this term indicates that they exhibit ergative case marking on the agent of a transitive verb. While most of these languages are like Margany, in that the ergative marking is found only on nouns and is complemented by accusative marking on pronouns, a handful of Pama-Nyungan languages – including Warlpiri, Kalkutungu (Kalkutung), and Yalarnnga – have ergative marking on both nouns and pronouns in A function, but no accusative marking on any free nominals. About two-thirds or more of Australian languages have bound pronominal representation for core functions, and these bound pronouns, with only a very few partial exceptions, operate on the basis of a subject (S and A) form and an object form (O).

Dixon (1972) argued that in Dyirbal, syntactic rules are sensitive to the grouping S + O, as opposed to A. This phenomenon has come to be referred to as ergative syntax, as opposed to accusative syntax; the latter term refers to a system of syntactic rules based on the notion of S + A (i.e., subject), as in English and numerous other languages. Ergative syntax is also found in some of Dyirbal's neighbors, including Yidiny, and in two adjacent languages of western Queensland, Kalkutungu (Kalkutung) and Yalarnnga. It manifests itself in a number of rules. For instance, there is a requirement that in relative clauses, the relativized function, which is covert, can be only S or O. To relativize an agent, the relative clause must be detransitivized via the antipassive, which thereby converts a potential A into S. In purpose clauses (also used for indirect commands), antipassive is used to signal that A is coreferent with S or O. The examples in (12) and (13) are from Yalarnnga. In the nature of things, coreference between S and A is common (as in [12]) and between P and A (as in [13]). In both these patterns of coreference, the antipassive is used.

- (12) Ngani-mi ngiya manhi-wu miya-li-ntjata.
go-FU I food-DAT get-AP-PURP
 'I'll go and get food.'
 (13) Tjuwa tjala ngathu ngapa-mu,
boy this I.ERG tell-PT

watjani-wu pinpa-li-ntjata.
wood-DAT gather-AP-PURP
 'I told this boy to gather firewood.'

The example in (14) provides a nice contrast. Here there is coreference between S in the second clause and P in the third, and there is no antipassive.

- (14) Ngathu tjala ngapa-mu ngani-ntjata
I.ERG this tell-PT go-PURP
 marnu-yantja-mpa karri-ntjata.
mother-HIS-ALL wash-PURP
 'I told him to go to his mother and get washed.'

In (15) there is coreference between A and A, and no antipassive.

- (15) Ngathu miya-ntjata yimarta
I.ERG get-PURP fish
 yunkunhi-nti-yarta yita-wampa.
return-CAUS-PURP this-ALL
 'I am going to get some fish and bring it back here.'

It appears, however, that ergative syntax is not common in Australia, despite the widespread use of ergative case marking. In a number of languages with ergative case, there are syntactic rules based on the familiar grammatical relation of subject (S + A). Many such rules have to do with showing maintenance of reference or switch reference. In Pitjantjatjara, for instance, the conjunction *munu* is used to link clauses with the same subject (SS), while *ka* is used to link clauses where there is a change of subject (DS). The point is that the rules operate on the basis of S and A, not S and O, as in languages like Dyirbal.

- (16) Tjitji panya ngarrikati-ngu munu
child that lie-PT and.ss
 ngarri-ngi kunkunpa ka kurta
lie-PT-IMPF sleep and.DS old.bro
 panya paka-rnu.
that get.up-PT
 'The child lay down and was lying asleep and the older brother got up.'

In a small group of Pama-Nyungan languages in western Australia, there is no ergative marking at all. The subject (S + A) appears in the nominative case, and the object (O) in the accusative/dative case. This group includes Ngarluma, Panyjima (Panytyima), and Yindjibarndi. It has been argued that these accusative languages derive from ergative languages via the generalizing of detransitivized constructions of the type illustrated in (10) and (11).

The non-Pama-Nyungan or Northern languages span the Northern part of the continent from western Australia to the Gulf of Carpentaria. With a few exceptions, mostly at the eastern end of their range,

the Northern languages have bound pronominal elements for subject and object prefixed to the verb. In some languages, these pronominal elements are separable, but more often than not, they fuse to one another and to other formatives in the verb. There is no accusative case marking on nouns or free pronouns, though there is ergative marking in some languages.

Among the prefixing languages, but also in the northwestern suffixing languages, it is common to find that only certain verbs can bear inflection. These verbs can appear on their own, as in (17), or they can act as auxiliaries in concert with an uninflected lexical verb, as in (18). These examples are from Maranungku (Maranunggu).

- (17) Tawun kangani yi.
town NONFU.I.go PT
 'I went to town.'
- (18) Tirr wuttar wat kangani yi.
edge sea walk NONFU.I.go PT
 'I walked to the beach.'

Systems of noun classes are common among the Northern languages, though a rarity in the Pama-Nyungan family. A majority of the Northern languages of the Kimberleys and the Top End have from two to eight noun classes, with each class marked by a prefix. The classification typically includes a masculine class, a feminine class, and a class for vegetable food. It is thought that these class markers are derived from generic nouns. It is not uncommon in Australian languages to use a generic noun accompanied by a specific noun. See, for instance, (26). The vegetable class marker is *m-*, *ma-* or *mi-*, and *mayi* is a widespread word for 'vegetable food,' so it is thought likely that the former derives from the latter. These class markers may appear not only on nouns representing direct dependents of the verb but also on associated demonstratives and appositional nouns. They may also appear on the verb, where they serve as crossreferencing pronominal forms. In the following example from Ngandi, *ni* (masculine) and *gu* (marking one of the inanimate classes) appear prefixed to the subject and object nouns respectively, and they are also prefixed to the verb.

- (19) Ni-gu-may ni-yul-thu gu-dyundu.
NI-GU-got NI-man-ERG GU-stone
 'The man got the stone.'

The noun phrases in (19) can be omitted. *Ni-gu-may* can stand as a sentence on its own, meaning 'He got it' or, more precisely, 'A member of the *ni* class got a member of the *gu* class'.

A feature of the prefixed bound pronoun systems is the prevalence of hierarchical principles of ordering

or marking. In Gunwinygu (Gunwinggu) (Northern Territory), first and second person forms always precede third, irrespective of which is subject.

- (20) Nga-n-di bun.
 1SG-OBJ-3PL hit
 ‘They hit me.’

- (21) Nga-be-n bun.
 1SG-3PL-OBJ hit
 ‘I hit them.’

The form *n-* glossed as object is common among the prefixing languages. In some languages this behaves like an inverse marker, in that it is used only when a person lower on the hierarchy acts on a higher person. This is the situation in Rembarnga (Rembarunga), where the hierarchy is 1 > 2 > 3PL > 3SG. Note that it does not appear in (22), where first acts on third, but it does appear in (23), where third plural acts on first.

- (22) Pa-nga-na.
 3PL-1SG-saw
 ‘I saw them.’

- (23) Nga-n-pa-na.
 1SG-O-3PL-saw
 ‘They saw me.’

A number of Northern languages incorporate nominals into the verb. The incorporated forms are often different from the corresponding words used outside the verb, and the range of concepts that can be incorporated is usually relatively small. The following example is from Tiwi, in which the incorporated form *wuliyondyi* refers to the direct object represented by *ti*.

- (24) Pi-ti-wuliyondji-rrurlimpirr-ani.
 3PL-3SG.FEM-dead.wallaby-carry.on.shoulders-
 PT.HABIT
 ‘They would carry the wallaby on their
 shoulders.’

Incorporated forms tend to correspond to the object of the verb, but they can correspond to other complements or adjuncts or to the subject of an intransitive verb.

Example (24) is fairly typical of Tiwi and of a number of other Northern languages that can be described as polysynthetic incorporating languages. Tiwi is obviously of quite a different type from Margany, which has no bound pronouns, or even Pitjantjatjara, which does. Tiwi has no case marking at all, and relations of complements and adjuncts to the verb are signaled via three series of bound pronouns representing subject, direct object, and indirect object, plus a few local prepositions. Not only are relations within the clause

marked on the verb, but the possessive relation is signaled within phrases by cross-referencing the possessor on the possessed (head) noun. ‘Purrukuparli’s son’ is expressed as *Purrukuparli ngarra-mirani*, literally ‘Purrukuparli, his son.’

Most Australian languages appear to have very free word order. Not only can the predicate, its complements, and its adjuncts appear in any order, but even the sets of words that translate a noun phrase of English may be separated. A common pattern is for a more general term, such as a pronoun or a generic noun, to be placed first, with the modifier late, often at the end. The example in (25) is from Nyangumarda (Nyangumarta).

- (25) Nyungu ngawu tjininganinyi
 this mad make.1PL.O
 walypila-mila-lu kari-lu.
 white.man-GEN-ERG beer-ERG
 ‘This is making us silly, the white man’s beer.’

The strategy employed in (25) is common in Australian languages. Another variation on this tendency is to use a generic noun early in the sentence and then a specific noun later, as in (26), from Yidiny.

- (26) Ngayu minya bugang ganguul.
 I animal eat wallaby
 ‘I’m eating wallaby.’

The fact is that most Australian languages have pragmatic principles rather than grammatical rules for word order. One such principle that is widespread is to put the focus (the emphasized phrase) first. There is probably no Australian language with word order as rigid as in English, but some languages have very strongly preferred orders. Some languages in the interior of the continent, including Pitjantjatjara, have fairly regular subject-object-verb order, and a few, such as Garrrwa (Garawa), are predicate-initial.

Semantics

The Australian Aborigines were hunter-gatherers, and naturally the vocabularies of Australian languages are rich in terms for fauna and flora as well as in terms for hunting and catching animals. There is regularly a distinction, for instance, between hitting or killing with a missile and hitting or killing with the hand or a handheld implement. There are words for decoy devices for attracting birds, words for a noose on a stick to catch a bird, words for different kinds of spears and boomerangs, and so on. Some semantic distinctions that are quite different from any made in European languages intrude into the grammar. In some Northern languages, there are forms for ‘you and I’ that pattern as singulars, i.e., the speaker and

Table 2 Gunwinygu (Gunwinggu) pronominal prefixes

| Number | Singular | Dual | Plural |
|--------|---------------|---------------|----------------|
| 1 | <i>nga-</i> | <i>ngane-</i> | <i>ngarri-</i> |
| 12 | <i>ngarr-</i> | <i>kane-</i> | <i>karri-</i> |
| 2 | <i>yi-</i> | <i>ngune-</i> | <i>ngurri-</i> |
| 3 | Ø | <i>bene-</i> | <i>birri-</i> |

addressee are treated as a unit. This becomes obvious when we examine the distribution of dual and plural marking. In **Table 2** the prefixed pronominal forms in Gunwinygu (Gunwinggu) are presented. The form *ngarr-* for ‘thou and I’ does not take dual marking but contrasts with a dual-marked form *ka-ne*, meaning ‘speaker and two addressees,’ and a plural-marked form *ka-rrri*, meaning ‘speaker and three or more addressees.’

In some languages there are different nonsingular pronouns for the kinship relations between the people referred to. In Alyawarra (Alyawarr), for instance, *mpula* means ‘you two’ but is used only for two people who belong to the same section. Among the Alyawarra everyone belongs to a patrimoiety, and within each patrimoiety there are two sections of alternating generations. There is a separate pronoun, *mpulaka*, for two people who are members of the same patrimoiety but not the same section (e.g., father and child), and a third form, *mpulantha*, for two people belonging to different patrimoieties (e.g., mother and child). This system of distinctions applies to all dual and plural pronouns.

Avoidance and Secret Vocabularies

In Aboriginal society it is common to have a special vocabulary that is to be used in the presence of certain kin. Normally a man is required to avoid dealings with his mother-in-law, for instance, and the prohibition covers real, prospective, and classificatory mothers-in-law. In some areas a man is required to use the special vocabulary in the presence of a mother-in-law, and such special vocabularies have come to be called ‘mother-in-law languages,’ though they are not separate languages, nor are they always reserved for speech in the presence of a mother-in-law.

Secret languages have also been reported from a number of areas. Like forms of avoidance language, these are special vocabularies usually taught as part of male initiation.

All these special vocabularies are of great linguistic interest. They typically consist of only a few hundred words, and often one finds a generic term in the reduced vocabulary that is lacking in the everyday language. In the avoidance language of the

Dyirbal (Queensland), for instance, there is a single word, *dyidyan*, for any lizard, skink, or goanna, and a single word, *dyiburray*, for any possum, squirrel, or glider. However, in the everyday language, there are words only for particular species (Dixon, 1980: 61).

Sign Language

Over much of central and northern Australia, sign language is used as an alternative to speech. Signs are made with the hands and correspond to words in the spoken language and to particles and suffixes that have local meanings, such as ‘to’ or ‘here.’ Sign language is traditionally used in a variety of contexts, including rituals, during periods of mourning when speech is proscribed, in conversing over long distances, or in hunting, where silence is important.

The Future of Australian Languages

Only a score or so of Australia’s native languages are being passed on to the next generation. Over the past three decades, there have been bilingual programs aimed at helping Aboriginal languages survive, and there is at least one instance of a language’s being revived, namely Dyaabugay in Queensland. There are also attempts at reclamation of languages no longer spoken, but the materials available for many languages, particularly in the southeast of the continent, are inadequate, and the best that future generations can hope for is to learn about their languages rather than acquire their languages. Some languages that are still spoken are undergoing drastic changes. Modern Tiwi, for instance, is much more analytic than traditional Tiwi, which is polysynthetic. For many Aborigines in the north of the continent, a creole is the first language – Torres Strait Broken (Torres Strait Creole), for example, spoken on Cape York, or Kriol in the Kimberleys and the Northern Territory. These creoles have a lexicon largely from English, with an admixture of vernacular vocabulary. They have some claims to being Aboriginal languages, not only on the grounds that they serve to mark Aboriginal identity but also in that they embody traditional semantic concepts that are calques from the vernaculars.

For most Australians, Aboriginal languages are a closed book, though there is a testimony to their existence in a few hundred words borrowed from Aboriginal languages – including *kangaroo* (Guugu-Yimidhirr, Guguyimidjir), *boomerang* (Dharuk), and *dingo* (Dharuk) – and thousands of place names, including Geelong (*tjilang* ‘tongue’), Warrnambool (*warnam-bul* ‘having fire’), and Wagga-Wagga (*waga-waga* ‘crows’).

Bibliography

- Blake B J (1988). 'Redefining Pama-Nyungan: towards the prehistory of Australian languages.' *Aboriginal Linguistics* 1, 1–90.
- Bowern C & Koch H (eds.) (2004). *Australian languages: classification and the comparative method*. Amsterdam: Benjamins.
- Capell A (1956). *A new approach to Australian linguistics*. Oceania Linguistic Monographs No. 1. Sydney: University of Sydney.
- Dixon R M W (1972). *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.
- Dixon R M W (1980). *The languages of Australia*. Cambridge: Cambridge University Press.
- Dixon R M W (2002). *Australian languages: their nature and development*. Cambridge: Cambridge University Press.
- Dixon R M W & Blake B J (eds.) (1979, 1981, 1983, 1991, 2000). *The handbook of Australian languages*, vols 1–3. Canberra: ANU Press; Amsterdam: John Benjamins. Vols 4–5, Oxford: Oxford University Press.
- Dixon R M W, Ramson W S & Thomas M (1990). *Australian Aboriginal words in English*. Melbourne: Oxford University Press.
- Evans N (1988). 'Arguments for Pama-Nyungan as a genetic subgroup, with particular reference to initial laminalization.' *Aboriginal Linguistics* 1, 91–110.
- O'Grady G N, Voegelin C F & Voegelin F M (1966a). 'Languages of the world: Indo-Pacific fascicle six.' *Anthropological Linguistics* 8, 1–197.
- O'Grady G N, Wurm S & Hale K (1966b). *Map of Aboriginal languages of Australia: a preliminary classification*. Victoria: University of Victoria.
- Silverstein M J (1976). 'Hierarchy of features and ergativity.' In Dixon R M W (ed.) *Grammatical categories in Australian languages*. Canberra: A I A S; New Jersey: Humanities Press.

Austric Hypothesis

L A Reid, University of Hawai'i, Honolulu, HI, USA

© 2006 Elsevier Ltd. All rights reserved.

Austric is the name given by the German missionary priest Wilhelm Schmidt (1906: 81–82) to the hypothesis that the Austronesian and Austro-Asiatic language families (also first named by Schmidt) are genetically related. Other versions of the hypothesis either include or exclude the Tai-Kadai language family of Southeast Asia, and/or the Hmong-Mien language family. The term 'Macro-Austric' is sometimes applied to a phylum which includes the Hmong-Mien language family.

The possible relationship of an Austro-Asiatic language, Nicobarese, with languages in what was then known as Malayo-Polynesian was first proposed in the latter part of the 19th century, but it was Schmidt who made the first systematic comparison of the two families, citing a considerable number of lexical comparisons, and claiming "complete agreement in phonology, morphology and various features of the syntax." Most of the lexical similarities cited by Schmidt have since been rejected by linguists (Diffloth, 1994) as not being adequately supported by regular sound correspondences. Nevertheless, the search for possible lexical cognates between the two language families continues. The most ambitious work in recent times has been that of Hayes

(1999, and earlier works). However a careful review (Reid, 2004) of Hayes's proposed basic vocabulary comparisons revealed that only a very small percentage are probable cognates supported by the usual requirements of regular sound correspondences and semantic similarity.

As for their phonology, morphology, and syntax, it is clear from the extensive descriptive materials that have been published since Schmidt's time that there is certainly not the "complete agreement" that Schmidt claimed for them. However, there are a number of puzzling similarities which call for explanation, especially when Nicobarese is considered. The aspect of Nicobarese that first stimulated Schmidt and others to note its similarities to Austronesian was not only that the language was typologically similar to languages such as Malay (with which they usually compared it) in having prefixes, infixes, and suffixes attached to verbs, but that the form and function of these affixes in many respects appeared to be similar to those in many Austronesian languages. Some of these features were first discussed by Schmidt (1916), and were expanded on in Reid (1994, 1999). Much of the following discussion is based on these two papers.

Typologically, Nicobarese is unlike other Austro-Asiatic languages in being a verb-initial language. In many respects it appears to be an Austronesian language with Austro-Asiatic lexicon. It has been generally characterized as SVO (Schmidt, 1906); however, text materials show numerous examples

of VOS word order, found for example in Tagalog, Malagasy, and other Austronesian languages. Noun phrase structure in Nicobarese is also strikingly similar to that found in many Austronesian languages, with noun-attribute word order, and attributes such as relative clauses linked to their head nouns with a form *na*, which commonly occurs in Austronesian languages with identical function. The same form also links adverbial attributes to their head verbs, just as in Austronesian. Noun phrases are introduced by one of a set of distinct case-marking forms, some of which have identical shape and function with those found in Austronesian languages. In morphology, there are a number of affixes, such as the causative prefix *ha-* (from earlier **pa-*), the agentive affixes *<um>* and *ma-l <am>*, the nominalizing infixes *<an>* and *<in>*, and the objective suffix *-a*, which are taken to be cognate with Austronesian affixes with the same or similar shape, and similar if not identical functions.

The main alternative explanation that has been proposed by those who reject a genetic relationship to account for these facts is borrowing. The claim has been made that the morphosyntactic features found in Nicobarese that appear to be Austronesian are probably remnants of a language spoken by early Austronesian sailors who may have made frequent landfall in the Nicobars, perhaps in some cases staying, intermarrying, and influencing the local language. But there remain several strong barriers to acceptance of this position. One is that several of the proposed comparisons between Nicobarese languages and Austronesian are not limited to Nicobarese, but are found across wide areas of the Austro-Asiatic family. In some cases (especially *<um>* and *<in>*), comparisons are clearest between Nicobarese and Austronesian, because other eastern Austro-Asiatic languages have either lost the form (in the case of verbal suffixes) or modified them due to the strong areal influence of Chinese. Another argument against the borrowing scenario is that some of the forms that are apparently of Austronesian origin predate Proto-Malayic and had changed by the time Austronesian sailors could have reached the Nicobars. A third argument against the borrowing hypothesis that has been proposed is that it is highly unlikely that a language could borrow so much morphology without also borrowing any of the lexical forms which carried it.

The only other possible explanation, according to Reid (1994), is a genetic one. The claim is that Nicobarese is a very conservative Austro-Asiatic language, a classic example of a 'relic' language because of its geographic isolation, lying far off the coast of mainland Southeast Asia, uninfluenced by

the leveling influences of Chinese and subsequently Thai that have produced the set of areal features commonly found in Mon-Khmer and other Austro-Asiatic languages. Nicobarese therefore is considered to reflect much of what must be reconstructed for the morphology and syntax of Proto-Austro-Asiatic and ultimately Proto-Austic.

Despite the lack of verifiable lexical comparisons and sustainable sound correspondence sets, some linguists still believe the Austic hypothesis has merit, considering the fairly substantial body of morphosyntactic evidence outlined above. Blust (1996) even proposes a homeland for Proto-Austic, in the general area of the watersheds of the Salween, Mekong and Yangtze rivers in the upper Burma-Yunnan border area. He claims that pre-Austronesians separated from this homeland around 7000 B.C., gradually moving down the Yangtze River valley till they reached the coast, and eventually sailed south and across the Taiwan Strait to Formosa. These proposals, however, have not been widely accepted.

The most recent challenge to the Austic hypothesis has come from Sagart (2004), who proposes an alternative genetic relationship for Austronesian. He claims that Austronesian is most closely related to Sino-Tibetan, and that at least some of the morphological features that appear to support the Austic hypothesis were present also in the parent of Sino-Tibetan-Austronesian, and therefore possibly give evidence of a relationship with Austro-Asiatic at a much greater time depth.

Bibliography

- Blust R (1996). 'Beyond the Austronesian homeland: the Austic hypothesis and its implications for archeology.' *Transactions of the American Philosophical Society* 86(5), 117–140.
- Diffloth G (1994). 'The lexical evidence for Austic, so far.' *Oceanic Linguistics* 33, 309–322.
- Hayes L V H (1999). 'On the track of Austic, part 3: basic vocabulary comparison.' *Mon-Khmer Studies* 29, 1–34.
- Reid L A (1994). 'Morphological evidence for Austic.' *Oceanic Linguistics* 33(2), 323–343.
- Reid L A (1999). 'New linguistic evidence for the Austic hypothesis.' In Zeitoun E & Li P J-K (eds.) *Selected papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academia Sinica. 5–30.
- Reid L A (2005). 'The current status of Austic: a review and evaluation of the lexical and morphosyntactic evidence.' In Sagart *et al.* (eds.). 132–160.
- Sagart L (2005). 'Sino-Tibetan-Austronesian: an updated and improved argument.' In Sagart *et al.* (eds.). 163–178.
- Sagart L, Blench R & Sanchez-Mazas A (eds.) (2005). *The peopling of East Asia: putting together archaeology,*

linguistics and genetics. London/New York: Routledge-Curzon.

Schmidt W (1906). 'Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens.' *Archiv der Anthropologie, Braunschweig n. s.* 5, 59–109.

Schmidt W (1916). 'Einiges über das Infix mn und dessen stellvertreter p in den austroasiatischen Sprachen.' In *Aufsätze zur Kultur- und Sprachgeschichte, vornehmlich des Orients, Ernst Kuhn zum 70*. Breslau: Marcus. 457–474.

Austroasiatic Languages

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

The Austroasiatic languages are spoken in small, often remote and inaccessible, hilly or mountainous regions throughout Southeast Asia, as far west as central India and as far east as Vietnam. There are over 150 languages belonging to the numerous Austroasiatic subgroups, enumerated below.

The primary split in the family is between the Munda languages in central and eastern India and the rest of the family. While lexically it is clear that Munda belongs to Austroasiatic, structurally the highly synthetic Munda languages are radically different from their predominantly isolating sister languages to the east. There are two major Munda subgroups, North Munda and South Munda (*see Munda Languages*).

Nahali (Nihali), an enigmatic group who speak a language that may or may not belong to Austroasiatic, are now mostly living as subjects to the North Munda Korku in the Indian states of Madhya Pradesh and Maharashtra. Some consider Nahali to have a special relation to Munda, others consider it to be a separate but related group of Austroasiatic, a third faction consider Nahali to be an isolated group in South Asia, like Burushaski (*see Burushaski*), while a fourth group of researchers reject Nahali as an independent language, rather considering it to be some kind of thieves' argot or secret language. Exact numbers of speakers are hard to gauge but may be around 5000.

There are at least three other major subgroups of Austroasiatic, the internal relations of which are still a subject of dispute. One such group is Nicobarese, which consists of a small number of languages spoken in the various Nicobar Islands, which lie off the southeastern coast of India, to which they belong administratively. Among this group of languages, Car Nicobarese, Nancowry Nicobarese or Central Nicobarese have received the most amount of linguistic investigation. One language, Shompeng (Shom Peng), appears to be highly divergent within the group,

but the materials on this language remain scanty. Other Nicobarese languages include Southern Nicobarese, Chaura (Chowra), and Teresa. Published sources include Radakrishnan's (1981) study of Nancowry morphology, among others. The total number of speakers of all Nicobarese languages is likely less than 25 000.

The next major subgroup of Austroasiatic is the Aslian group, which is spoken primarily in Malaysia (where the speakers are known as Orang Asli) but also in adjacent areas of Thailand. Ethnoracially, the Orang Asli of Malaysia fall into three subgroups: the Semang/Negrito, the Sakai/Senoi, and the Jakun/Aboriginal Malay (Parkin, 1991: 41). The first option in each case was traditional but has now become stigmatized, and the latter variant is now preferred. (Note that curiously the Semang/Negrito speakers prefer Sakai, although this is considered offensive to those whom it originally designated; cf. Parkin, 1991: 42.) Only two Jakun/Aboriginal Malay groups speak Aslian languages, Semelai and Temoq. Importantly, the linguistic subgroups of Aslian do not correspond neatly (although partially) to this ethnoracial categorization. In particular, there appears to be a primary split between a southern group (Semelaic (Semelai)) a northern, and a central subgroup (Jahaic (Jehai) and Senoic, respectively). Jah Hut may constitute an isolate branch within Aslian, although others consider it a divergent member of the Senoic subgroup. The exact relation between these subgroups remains to be worked out explicitly. Jahaic includes Negrito groups as well as racially Senoic Chewong. Jahaic languages are mainly spoken by very small groups of a few hundred speakers at most. None could be described as well known, but the subgroup includes such languages as Kintaq (Kintaq Bong), Minriq, Mintil, Jehai (Jahai), Batek, Tonga/Mos, which is mainly spoken in Thailand, Kensiu, and probably the Lowland Semang of Sumatra, with nearly 10 000 speakers. Senoic languages consist of several subgroups. The most important of these are the Lanoh, the poorly known Sabüm, the Temiar, and especially the Semai, who are the largest Aslian-speaking group with possibly as many as 20 000

speakers. Temiar, with perhaps 10 000 speakers, has been an important loan source for Jahaic languages, and is one of the best-studied members of this group (Carey, 1961; Benjamin, 1976). The Semelaic (South Aslian) branch consists of a small number of languages, each of which has probably fewer than 2000 speakers. In addition to Semelai and Temoq, the languages include Semaq Beri and Maq Betiseq (Besisi), also known as Mah Meri. Semelai has recently become the best studied of all Aslian languages with the publication of a large grammar by Nicole Kruspe (2004).

The fourth and final major subgroup within Austroasiatic is the far-flung Mon-Khmer group. This has a number of different subgroups, the internal relations of which remain to be adequately worked out. Major languages in this subgroup include Khmer (Cambodian, Khmeric), Mon (Monic), Vietnamese (Viet-Muong), Khasi (Khasic), Bahnar (Bahnaric) [BDQ], Kuy (Katuic), Palaung (Palaung-Wa), including Pale, Rumi, and Shwe, and so forth (*see Mon-Khmer Languages*).

Generally speaking, the westernmost languages of the family exhibit the greatest degree of morphological development. Munda languages are inflectional and agglutinating, with a diverse and highly developed system of tense/aspect marking, subject and object agreement, noun incorporation, and so on. An extreme example of this comes from Kharia, where the following word has no fewer than 8 morphemes:

- (1) *Kharia*
ḍoḍ-kay-tu-dom-bhaḷ-god-na-m
 carry-BEN-TLOC-PASS-quickly-COMPLT-FUT-2
 'get yourself there for me quickly'
 (Malhotra, 1982)

Tense/aspect morphology is not common among non-Munda Austroasiatic languages but may be found in Lyngngam of the Khasic branch of Mon-Khmer (*see Khasi*) and in certain Bahnaric and Katuic languages. In addition to Munda, certain Aslian languages show subject agreement in the verb, but otherwise this feature is not a common one in Austroasiatic.

South Munda and Nicobarese, and to a lesser extent the Aslian language Temiar, reflect evidence of noun incorporation, and this may therefore have been a feature of earlier stages of the Austroasiatic language family.

- (2) *Temiar*
pasal-naq ki-chiibjuq
 reason-that 1PL-walk < *'go.foot'
 'so we had to go on foot'
 (Carey, 1961: 46)

It seems certain that Proto-Austroasiatic was richer morphologically than the majority of Mon-Khmer languages, particularly in terms of derivation, but not as developed as the Munda languages. Among the more noteworthy features of Austroasiatic is the unusually frequent use of infixation processes. A small number of derivational elements appear to be cognate across the members of the family, for example, a causative verb formant and a nominalizing element. The former appears either as a prefix or an infix, depending on the stem shape. Both elements are found in such branches as Munda, here exemplified by Juang Nicobarese, and the Mon-Khmer subgroups Monic and Khmuic (Khmuic), while other branches preserve only the prefix allomorph.

- | | |
|---|---|
| (3) <i>Juang</i> <i>a'b-soŋ</i> CAUS-buy 'sell' (Pinnow, 1960a) < <i>kəsoŋ</i> | <i>Juang</i> <i>kə'b-səŋ</i> dry..-CAUS-..dry 'dry sthg' |
| (4) <i>Nancowry</i> <i>ha-kab-naŋ</i> CAUS-know-ear 'make understand' (Radakrishnan, 1981: 87) | <i>Nancowry</i> <i>p-um-lōŋ</i> lose-CAUS-lose 'make lose' (Radakrishnan, 1981: 54) < <i>plōŋ</i> |

Another infixation process found across the languages of the Austroasiatic stock is the nominalizing infix -n-. This is found in such forms as Khasi *shmong* 'village' < *shong* 'live,' Mlabri *chnræt* 'comb' < *chræt* 'to comb,' or Mundari *dunub* 'meeting' < *dub* 'sit.'

It has been put forth that Austroasiatic may be a part of a larger genetic unit. Various proposals include relations with Austronesian, Tai-Kadai, Hmong-Mien (Miao-Yao), and even Sino-Tibetan, variously labeled 'Austic,' Austro-Tai, and so on. None of these proposals are widely accepted by specialists, and these hypotheses should therefore be treated with caution. Among modern specialists in Austroasiatic languages, Gerard Diffloth deserves special mention.

Bibliography

- Adams K (1989). *Systems of numeral classification in the Mon-Khmer, Nicobarese, and Aslian subfamilies of Austroasiatic*. Canberra: Australian National University.
 Anderson G D S & Zide N H (2001). 'Recent advances in the reconstruction of the proto-Munda verb.' In Brinton L (ed.) *Historical linguistics 1999*. Amsterdam: Benjamins. 13–30.

- Benedict P K (1976). 'Austro-Thai and Austroasiatic.' In Jenner *et al.* (eds.), 1–36.
- Benjamin G (1976). 'An outline of Temiar grammar.' In Jenner *et al.* (eds.), 129–188.
- Carey I (1961). *Tengleq Kui Serok*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Das A R (1977). *A study on the Nicobarese languages*. Calcutta: Anthropological Survey of India.
- Diffloth G (1976a). 'Minor-syllable vocalism in Sinoic languages.' In Jenner *et al.* (eds.), 229–248.
- Diffloth G (1976b). 'Jah Hut, an Austroasiatic language of Malaysia.' In Nguyễn D L (ed.) *Southeast Asian linguistic studies 2. Pacific linguistics C-42*. Canberra: Australian National University. 73–118.
- Diffloth G (1989). 'Proto-Austroasiatic creaky voice.' *Mon-Khmer Studies* XV, 139–154.
- Ferlus M (1974). 'Les langues du groupe Austroasiatique-Nord.' *Asie du Sud-Est et Monde Insulindien* 5(1), 39–68.
- Haji Omar A (1976). 'The verb in Kentakbong.' In Jenner *et al.* (eds.), 951–970.
- Haudricourt A-G (1954). 'De l'origine des tons en vietnamien.' *Journal Asiatique* 242, 69–82.
- Jenner P N, Thompson L C & Starosta S (eds.) (1976). *Austroasiatic studies I*. Honolulu: University of Hawaii Press. [Oceanic Linguistics Special Publication No. 13.]
- Kruspe N (2004). *A grammar of Semelai*. Cambridge: Cambridge University Press.
- Kuiper F B J (1966). 'The sources of Nahali vocabulary.' In Zide N (ed.) *Studies in comparative Austroasiatic linguistics*. The Hague: Mouton. 57–81. [Indo-Iranian Monographs 5.]
- Malhotra V (1982). The structure of Kharia: a study in linguistic typology and change. Ph.D. diss., J Nehru University.
- Parkin R (1991). *A guide to Austroasiatic speakers and their languages*. Honolulu: University of Hawaii Press. [Oceanic Linguistics Special Publication 23.]
- Pinnow H-J (1959). *Versuch einer Lautlehre der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.
- Pinnow H-J (1960a). *Beiträge zur Kenntnis der Juang-Sprache*. Unpublished manuscript, Berlin.
- Pinnow H-J (1960b). 'Über den Ursprung der voneinander abweichenden Struktur der Munda und Khmer-Nikobar Sprachen.' *Indo-Iranian Journal* 4, 81–103.
- Pinnow H-J (1963). 'The position of the Munda languages within the Austroasiatic language family.' In Shorto H L (ed.) *Linguistic comparison in Southeast Asia and the Pacific*. London: SOAS. 140–152.
- Pinnow H-J (1965). 'Personal pronouns in the Austroasiatic languages: a historical study.' *Indo-Pacific Linguistic Studies* 1, 3–42.
- Radhakrishnan R (1981). *The Nancowry word*. Edmonton: Linguistic Research.
- Schmidt W (1901). 'Die Sprachen der Sakai und Semang auf Malacca und ihr Verhältnis zu den Mon-Khmer-Sprachen.' *Bijdragen tot de Taal-, Land-, en Volkenkunde van Nederlandsch-Indië* 52, 399–583.
- Shafer R (1952). 'Études sur l'austroasiatique.' *Bulletin de la Société de Linguistique* 48, 111–158.
- Temple R C (1902). *A grammar of the Nicobarese language. Being chapter IV of part II of the census report on the Andaman and Nicobar Islands*. Port Blair: Superintendent's Office.
- Zide N (ed.) (1966). *Studies in comparative Austroasiatic linguistics*. The Hague: Mouton. [Indo-Iranian Monographs 5.]

Austronesian Languages

T Crowley[†], University of Waikato, Hamilton, New Zealand

© 2006 Elsevier Ltd. All rights reserved.

The Language Family and Its Speakers

Austronesian is possibly the largest language family in the world. Its 1200 or so languages (Grimes *et al.*, 1994: 122) amount to about a fifth of the world's total number. While the Niger-Congo family is sometimes said to be larger than Austronesian by a couple of hundred languages, it is by no means a well-established grouping, and some have suggested that

it is merely a typological rather than a genetic grouping. The lower-level Benue-Congo grouping (see **Benue-Congo Languages**) is much better established, but it has about 200 fewer languages than the Austronesian family. Austronesian is way ahead of the next grouping, the Trans New Guinea languages, which has fewer than 600 members. However, Austronesian again constitutes a much more clearly recognizable family than the Trans New Guinea grouping.

Austronesian languages represent the fourth-largest grouping of languages in the world in terms of the number of speakers. According to some, they are beaten again by Niger-Congo languages, relegating them to fifth position, though the relatively poorly supported claims about the genetic unity of these languages means that the fourth position for

[†]Deceased.

Austronesian should perhaps be maintained. The total number of speakers of Austronesian languages is about 300 million, which represents about 5% of the world's population. The Austronesian family includes the world's 13th-largest individual language (Javanese) in terms of the number of native speakers. Malay/Indonesian (*see* **Malayo-Polynesian Languages**) and Tagalog (spoken in the Philippines) (*see* **Tagalog**) come in at 9th and 18th respectively in terms of the total number of first- and second-language speakers (Crystal, 1987: 287). No putative Niger-Congo language appears in the top 20 for either list.

If we exclude the spread of Indo-European languages to the New World in association with colonialism, Austronesian languages also have by far the largest geographical spread of any language family in the world. Their territory extends from the islands of Taiwan and Hawai'i in the north, Easter Island (or Rapanui) in the east, New Zealand in the south, and Madagascar in the west. However, the territory within these bounds is not occupied exclusively by speakers of Austronesian languages, as Australia (*see* **Australian Languages**) and Tasmania, parts of the New Guinea area, and parts of mainland Southeast Asia include a variety of different non-Austronesian languages.

The Austronesian family is noteworthy not just for its largest languages, as it includes a huge number of very small languages as well. The Republic of Vanuatu – located in the southwest Pacific – has a population of only about 200 000, but its people speak at least 80 separate Austronesian languages (Lynch and Crowley, 2001), giving each language an average population of about 2500 speakers and making Vanuatu possibly the world's most diverse nation in terms of the number of languages per capita.

While Austronesian languages constitute a well-defined linguistic grouping, their speakers are very diverse in terms of physical appearance. People of a variety of Asian types speak Austronesian languages in what is now Indonesia, Malaysia, Singapore, Brunei, the Philippines and the interior of Taiwan. In the far west in Madagascar, speakers of the Austronesian language Malagasy clearly exhibit African genes. In the Pacific, the Melanesian speakers of Austronesian languages from the island of Timor, the Indonesian province of West Papua, Papua New Guinea, Solomon Islands, Vanuatu, New Caledonia, and Fiji differ in appearance from their Asian neighbors to their west, from their Polynesian neighbors to their east, and from their Micronesian neighbors to their north.

In fact, the boundaries between physical types are far from rigid, and there is often a gradual transition from one type to another, as genes have been mixing for centuries. In any case, linguistic boundaries and the boundaries of physical types often fail to coincide. We see this most dramatically in the New Guinea area, where physically similar Melanesian peoples may speak Austronesian languages or any of a number of completely unrelated non-Austronesian languages, including languages belonging to the Trans New Guinea grouping referred to earlier in this article. Sometimes, people in neighboring villages may speak totally unrelated languages. In fact, in the agglomerated village of Hanuabada, in Papua New Guinea, speakers of Austronesian Motu and non-Austronesian Koita (Koitabu) live side by side in the same community.

The Austronesian-speaking area exhibits cultural diversity that is even more dramatic than the diversity of physical types. As an illustration, we could point to the Hindu culture of Bali, in Indonesia; the traditional animist belief systems of Austronesian speakers in Melanesia (which continue to be practiced in some areas); the traditional polytheistic practices of the Polynesians, the Muslims of most of Indonesia, Malaysia, and southern Philippines; and the centuries-old Christian traditions of the central and northern Philippines. In some parts of the Austronesian-speaking world, traditional culture areas may be only slightly larger than the areas occupied by some of the very small individual languages. For instance, on the island of Malakula, in Vanuatu, significant differences in social organization and material culture can be found over quite short distances, with distinct culture areas including only two or three quite small language groups.

Of course, there has been a great deal of relatively recent technological and cultural change throughout the Austronesian-speaking world, with the advent of European colonialism and the modern technological revolution. The changes have perhaps been most dramatic (and most recent) in Melanesia, where in some cases fully traditional practices held sway until the first half of the 20th century. Although there are unlikely to be any more dramatic discoveries of 'lost tribes' who know nothing of the outside world, there are certainly still places where contact has until now been fairly minimal. While Christianity has now been adopted with fervor in most of Melanesia, Micronesia, and Polynesia, this change has often taken place in a way that has allowed for the retention of various aspects of the traditional belief system along with (or as part of) local Christianity.

Internal Genetic Relationships and Reconstruction

Based on the analogy of the naming of the Indo-European family after the geographical extremities – Indian and European languages – the Austronesian languages were originally referred to as Malayo-Polynesian, after Malay (and its relatives) in the west and the Polynesian languages in the east. However, it was subsequently realized not only that the indigenous Formosan languages of Taiwan belonged in this family but also that these represented several distinct high-level subgroups. The term *Malayo-Polynesian* was then reassigned to cover all of the non-Formosan languages within the enlarged family. This was, henceforth, referred to as Austronesian, based on the elements *Austro-* ‘southern’ and *-nesia* ‘island.’ The latter element is, of course, also found in the names for the geographical areas of Polynesia (= many islands, because of the large number of islands involved), Melanesia (= black islands, because they are occupied by dark-skinned peoples), and Micronesia (= small islands, because these are mostly narrow, low-lying atolls).

While the Austronesian languages exhibit a considerable amount of structural diversity, the existence of the language family as a whole is completely uncontroversial, in contrast to that of some other language groupings – including the so-called Niger-Congo languages – where debates rage between ‘lumpers’ (who seek to link together as many languages on the basis of what sometimes look to others more like typological similarity) and ‘splitters’ (who are sometimes overcautious in requiring anything but absolutely infallible proof of genetic relationship).

In fact, the idea that Austronesian languages are related is often obvious even to a casual observer, in contrast, again, to many other language groupings where even an experienced linguist might find it difficult to see convincing evidence of a relationship. For example, the root for ‘eye’ is *mata* in exactly this shape in languages as far apart as Yami (in Taiwan), Tagalog (in the Philippines), Malay/Indonesian, Manggarai (in eastern Indonesia), Manam (in Papua New Guinea), Roviana (in Solomon Islands), Raga (in Vanuatu), and Tongan, Tahitian, and Rapa Nui (on Easter Island). Such lexical similarities are reasonably common in those parts of the vocabulary which we would expect to be most resistant to borrowing – hence strong indicators of genetic relationship – and borrowing is most unlikely, in any case, as an explanation for these similarities, given the huge distances involved.

So readily apparent is the relationship between many of these languages that a connection of sorts

between the Polynesian languages and Malay was suggested by Hadrian Reland as early as 1708, when very little indeed was known about most of these languages. Lorenzo Hervás y Panduro in 1784–1787 described a more detailed set of linguistic relationships among Austronesian languages in which the language of Madagascar and a larger number of Indonesian languages were also included (Lynch *et al.*, 2002: 1).

The little-known islands of Melanesia were usually excluded from these original generalizations, perhaps partly because of a mistaken assumption that the physically distinct Melanesian peoples should also be linguistically very distinct. However, it turned out from work in the late 19th century by H. C. von der Gabelentz and R. H. Codrington (Lynch *et al.*, 2002: 2), based largely on information supplied by Christian missionaries in the field, that a substantial number of these languages *do* belong in this family as well. While many of the languages of the New Guinea area are clearly *not* Austronesian, some of the Austronesian languages of Melanesia were originally thought to be non-Austronesian only because extensive phonological changes had obscured the shapes of many widely distributed Austronesian roots, or because extensive lexical innovation had led to the replacement of some of the more widespread Austronesian cognates by which relationship could be most easily recognized. By the end of the 19th century, however, it was realized that a substantial number of indisputably Austronesian languages were in fact spoken in many of the coastal parts of the New Guinea area, as well as in Solomon Islands, Vanuatu, and New Caledonia.

This burgeoning language family was soon to become the most serious early testing ground for the comparative method of phonological and lexical reconstruction that was developed initially on the basis of Indo-European – and, less widely known, Finno-Ugric – languages in the second half of the 19th century. Although Edward Sapir’s reconstruction of Uto-Aztecan in 1913–1915 represented a stunning early application of the comparative method to unwritten languages, Otto Dempwolff’s (1934, 1937, and 1938) comparative study of the vastly larger Austronesian family represented a much more challenging test of the method.

Of course, a large number of new Austronesian data have become available since Dempwolff’s time, and there has also been significant fine-tuning of the comparative method itself. Many of his comparisons were enriched in the work of Isidore Dyen (1951, 1953a, 1953b, 1965) and others from the 1950s, and many new reconstructions have also been proposed. Robert Blust (1970, 1980) has progressively

added to the reconstructed lexicon since the early 1970s, bringing the total number of lexical reconstructions now to many thousands of entries. A substantial amount of morphosyntactic reconstruction since the late 1970s can also be ascribed to the level of Proto-Austronesian as a result of work by Stanley Starosta (1985), Lawrence Reid (1978), and others.

Comparative reconstruction has not proceeded solely at the level of Proto-Austronesian, as there has been major effort devoted to lower levels of reconstruction as well. Perhaps the most significant intermediate reconstruction involves the ongoing work since the 1990s of Malcolm Ross *et al.* (1998, 2003) in the reconstruction of Proto-Oceanic, the ancestor of the 500 or so members of the Oceanic subgroup. However, many others have also contributed in this area, beginning with the work of George Grace (1969) and Wilhelm Milke (1968) in the 1960s. Below the level of Proto-Oceanic, there is a tradition of reconstruction of Proto-Polynesian, for which serious comparative work based on a wide selection of languages dates from David Walsh and Bruce Biggs (1966).

With such a huge language family, internal subgrouping could be expected to be a somewhat contentious issue. However, there is now broad agreement on many issues of subgrouping within Austronesian. Work on subgrouping methodology by Malcolm Ross since the 1980s has added new considerations to the subgrouping of Austronesian languages, with his distinction between separation-induced ‘subgroups’ on the one hand and ‘linkages’ that have arisen as a result of gradual diversification of dialects that remained in geographical contiguity (Ross, 1988). This distinction allows us to take into account the fact that some lower-level groupings of languages, rather than uniquely sharing a set of defining innovations, may actually overlap with neighboring groupings in that they may appear to share innovations from more than one subgroup.

While new data and fresh approaches to subgrouping methodology may bring about further revisions in the future, the generally accepted current view is that the area of greatest subgrouping diversity is on the island of Taiwan (Lynch *et al.*, 2002: 4). Recent research indicates that there may be as many as nine first-order subgroups there (Blust, 1999), with the remaining first-order subgrouping consisting of the Malayo-Polynesian languages, which is made up of the huge number of remaining Austronesian languages.

The western part of the Malayo-Polynesian subgroup appears to consist of a large number of smaller subgroups. This region includes all of the languages

of the Philippines, as well as Malaysia and the islands of Indonesia from Sulawesi and Sumbawa westward, and also the Malagasy language of Madagascar. It is in this area that all of the very large Austronesian languages belong, including Tagalog (*see Tagalog*), Sebuano, Ilokano (Ilocano), Hiligaynon, and Bikol (Bicolano) in the Philippines, and Malay/Indonesian, Javanese (*see Javanese*), Sundanese (Sunda), Madura, Minangkabau, Bugis, Balinese (Bali), and Acehnese (Aceh) in Indonesia.

All of the languages to the east of the Western Malayo-Polynesian languages probably belong in a single very large Central and Eastern Malayo-Polynesian subgrouping that consists overwhelmingly of much smaller languages. This subgroup is thought to involve a binary split between a geographically restricted Central Malayo-Polynesian grouping involving the languages of Sumba, Flores, Timor, Buru, Seram, and adjacent smaller islands, and a much larger Eastern Malayo-Polynesian grouping consisting of all the rest. However, the internal subgrouping of both Central and Eastern Malayo-Polynesian and Central Malayo-Polynesian remains poorly understood. Eastern Malayo-Polynesian in turn enters into a binary split between the Southwest Halmahera–West New Guinea languages on the one hand and the very large Oceanic subgroup on the other.

The Oceanic subgroup occupies a special place within Austronesian linguistics. Although this is by no means one of the highest-level subgroups in the family, it is nevertheless a huge grouping, comprising nearly half of all of the Austronesian languages and amounting to nearly 10% of all of the languages of the world. With a total Oceanic-speaking population of about 2 million, the average-sized Oceanic language can claim only about 4000 speakers. Excluding some of the largest Oceanic languages from this total, such as Fijian (*see Fijian*), with nearly 500 000 speakers, the average size for an Oceanic language drops to closer to 3000 speakers. Most of these languages are poorly documented in comparison to languages further west, and many are almost completely undocumented.

Oceanic subgrouping diversity is greatest in the west, with possibly four of the five primary subgroups located in this area: the Admiralties languages; the Western Oceanic languages of the north coast of the Indonesian province of Papua (formerly known as West Papua) and the coast of New Guinea, New Britain, New Ireland, Bougainville, and western Solomon Islands; the St Matthias subgroup; and the Yapese language as a single-language subgroup (Lynch *et al.*, 2002: 92–120).

A putative Central and Eastern Oceanic subgroup covers Polynesia and Fiji and all remaining areas of

Micronesia and Melanesia, including New Caledonia, Vanuatu, and part of Solomon Islands. Within this very large grouping, there is a five-way split between the Micronesian languages, the languages of south-eastern Solomon Islands, the languages of Utupua and Vanikoro in Solomon Islands, the languages of Vanuatu and New Caledonia, and finally the Central Pacific languages, consisting of Fijian, Rotuman, and the Polynesian languages.

Further conclusions have been presented about subgrouping at even lower levels, with a detailed subgrouping diagram available for the Polynesian languages. Given the size of the Oceanic family within Austronesian, the final family tree diagram is obviously going to be extremely complex. All subgrouping hypotheses will have to be kept 'open' pending further linguistic documentation in poorly known areas. One point that will be obvious, however, is that the geographically expansive and relatively well-described Polynesian languages, which have for more than 200 years figured so prominently in European fantasies about the Pacific, represent just a very small grouping of just over two dozen languages at the very lowest level of Austronesian subgrouping.

Linguistic Features

The huge size of the Austronesian family makes any kind of summary statements about 'typical' features well-nigh impossible. At the same time, the approximately 5000-year time depth for Austronesian languages is relatively shallow compared with language groupings such as Australian languages and Trans New Guinea languages, and the Austronesian family is structurally rather less diverse than such groupings as a result.

Since there are major structural differences between some of the Formosan and Western Malayo-Polynesian languages on the one hand and the Oceanic languages on the other, it is perhaps best to offer several sets of generalizations about widespread linguistic features in Austronesian languages. Even so, it must be recognized that within any such structural groupings, there are many languages that exhibit rather different sorts of patterns, so the patterns that are presented here are those which, in addition to having substantial geographical distribution, also appear to reflect some antiquity in a reconstructive sense.

In terms of phonology, it is particularly difficult to generalize about Austronesian languages. One thing that it is possible to say is that tonal contrasts are almost completely absent, in contrast to the phonological systems of many neighboring Asian

languages. Tone is furthermore not reconstructible at all in Proto-Austronesian. While there has long been general agreement on the reconstructed four-vowel system /i u ə a/, there has until recently been much less agreement firstly on the number of consonantal contrasts, and secondly on the precise phonetic value of a number of reconstructed protoconsonants.

Complex sets of consonantal correspondences led scholars in the past to posit consonant inventories varying between two and three dozen segments in total. Arguments were presented, for example, that correspondences pointed to a maximal set of reconstructions involving /d ɖ D D₁ D₂ D₃ d₁ d₂ d₃/, where the symbols represented a mix of phonetic and purely formulaic information (Ross, 1994: 54). These issues of phonetic uncertainty and the proliferation of protoconsonants have now been largely laid to rest (Blust, 1999), though Wolff (2003) remains a dissenting voice. Because the few Formosan languages represent a number of primary subgroups of Austronesian, the relatively recent study of these languages has provided evidence for substantial modifications in both the inventory and the phonetic value of Dempwolff's original phonological reconstructions.

There has been a considerable amount of phonemic merger, split, and shift in many subgroups and in many individual Austronesian languages. There has also been phonological erosion of particular phonotactic positions, particularly involving word-final consonants in Oceanic languages. Original nasal-stop clusters have also been reanalyzed in Oceanic languages as prenasalized stop phonemes, resulting in substantial phonotactic simplification. Some individual languages have undergone other more dramatic phonological changes, including also the reanalysis of other material as part of the root, so reconstructible **mata* 'eye' appears regularly as /nəmɾə/ in the Lenakel language of Vanuatu.

In terms of basic clause structure, the languages of the geographical extremities of the Austronesian-speaking areas are typologically very different. The languages of Taiwan and the Philippines exhibit what are often called 'focus' systems, which appear to directly reflect a reconstructible pattern at the Proto-Austronesian level (Ross, 1994: 64–66). In this system, verbs carry inflectional marking – expressed variously as prefixes, suffixes, or infixes or as a combination of more than one of these affixed elements – for so-called Actor Focus (AF), Undergoer Focus (UF), Locative Focus (LF), and Instrumental Focus (IF). The noun phrase that is signaled as being in focus typically appears clause-finally, is definite, and performs a range of possible semantic roles according to

the nature of the focus marking on the verb. This is a completely un-English construction for which it is difficult to provide close translations, but examples (1), (2), (3), and (4) from Tagalog illustrate this pattern, based on the verb roots *bili* ‘buy,’ *tamVn* ‘plant,’ and *putol* ‘cut’:

- (1) *B-um-ili ng kotse ang lalake.*
 AF:buy PATIENT car TOPIC man
 ‘The man bought a car.’

- (2) *B-in-ili ng lalake ang kotse.*
 UF:buy AGENT man TOPIC car
 ‘The car was bought by the man.’

- (3) *T-in-amm-an ng lalake ng damo ang lupa.*
 LF:plant AGENT man PATIENT
 grass TOPIC ground
 ‘The ground was planted the grass in by the man’ – i.e., ‘The man planted the grass in the ground.’

- (4) *I-p-in-utol ng lalake ng isda ang kutsilyo.*
 IF:cut AGENT man PATIENT
 fish TOPIC knife
 ‘The knife was cut the fish with by the man’ – i.e., ‘The man cut the fish with the knife.’

Oceanic languages, by way of contrast, evolved a system of formal marking for transitivity on verbs that expresses a range of different semantic roles associated with the verbal object. This marking involves verbal suffixes that express a distinction between unsuffixed transitive verbs with undergoer objects and those that carry a transitive suffix with typically oblique objects. These transitive suffixes are often described as expressing close and distant transitive respectively and are associated with different sorts of semantic roles. Thus, in Fijian the intransitive verb *qasi* ‘crawl’ corresponds to the close transitive form *qasi-va* ‘crawl to (location)’ and the distant transitive form *qasi-vaka* ‘crawl with (something held).’

Oceanic languages have evolved a wide range of other innovative features from the reconstructible Proto-Austronesian pattern. At the Proto-Oceanic stage, a formal distinction had developed between inalienable and alienable possession, with the former being expressed by means of pronominal suffixes attached directly to a noun and the latter expressed by means of an adposed possessive constituent to which pronominal suffixes were attached (Lynch *et al.*, 2002: 40–41). This distinction is still widely reflected in Oceanic languages in examples such as (5) and (6), from the Naman (Nama) language of Vanuatu:

- (5) *khavo-g*
 brother-1SG
 ‘my brother’

- (6) *neim khëso-g*
 house POSS-1SG
 ‘my house’

Naman has in fact simplified the reconstructible system for the expression of alienable possession, in that there is now only a single set of adposed possessive constituents. At the Proto-Oceanic stage, there are likely to have been separate possessive forms depending on whether the possessed item was for eating, for drinking, or for any miscellaneous purpose (Lynch *et al.*, 2002: 42). This three-way distinction is still maintained in many Oceanic languages, such as examples (7), (8), and (9), from Fijian:

- (7) *na me-mu wai*
 ART DRINK-2SG water
 ‘your water’

- (8) *na ke-mu dalo*
 ART ED-2SG taro
 ‘your taro’

- (9) *na no-mu vale*
 ART POSS-2SG house
 ‘your house’

In some Oceanic languages, however, a number of additional categories of alienable expression occur. In the Raga language of Vanuatu, for example, in addition to forms expressing edible, drinkable, and miscellaneous possession, there are the possessive constituents *bila-* ‘garden plots, crops, domestic animals, personal adornments’ and *wa-* ‘sugarcane,’ and some languages have developed even more categories of alienable possession.

The Polynesian languages have taken the expression of possession in yet another direction. A possessor is preceded by a possessive marker containing either the vowel *a* or *o*. Thus, in Samoan, we find examples such as those in (10) and (11) (Lynch *et al.*, 2002: 43):

- (10) *lo-‘u tama*
 POSS-1SG son
 ‘my son’

- (11) *la-‘u naifi*
 POSS-1SG knife
 ‘my knife’

Inalienably possessed items generally express possession by means of the *o* forms, while *a* forms tend to correspond to categories of alienable possession in other Oceanic languages. However, there is substantially more arbitrariness and metaphor involved in the interpretation of the alienable-inalienable distinction in these languages. For instance, the possession of canoes in Samoan is considered as being more ‘personal’ than the possession of a knife, so ‘my canoe’ is expressed as *lo-‘u paopao*.

Historical Interpretation

By combining subgrouping information, the content of the reconstructible lexicon of Proto-Austronesian (as well as the lexicons of various lower-level subgroups, most significantly Proto-Oceanic), and information provided by the archaeological record, it is possible to come up with a fairly sophisticated picture of Austronesian history dating back at least 5000 years (Bellwood *et al.*, 1995). The story of human settlement of the Pacific Islands has tended to attract the interest of enthusiastic amateurs and religious groups who have been prepared to argue particular points of view in a way that has sometimes attracted a certain amount of broader public acceptance, even though the scientific evidence often points in radically different directions.

Given that the area of greatest subgrouping diversity within a language family is most likely to represent the original homeland, we could argue on these grounds alone that the Austronesian homeland is likely to have been either on Taiwan or on the adjacent mainland of southern China and that there has been a general west-to-east movement, with Polynesia being the last area to be settled. Given that the Polynesian languages represent a linguistically very homogeneous and geographically expansive subgroup, this, on linguistic grounds, would be the least likely source for Austronesian languages. However, that has not prevented people inspired by the *Kon Tiki* expedition from arguing instead for a general east-to-west population movement, beginning in South America.

The view of some religious groups that Polynesian people are descendants of one of the Lost Tribes of Israel is also impossible to reconcile with the fact that there is no linguistic evidence in support of this contention, while there is a huge amount of incontrovertible evidence in support of linguistic relationships between Polynesian languages and those of the rest of the Austronesian-speaking world. And of course, there is a mass of archaeological evidence pointing to the origin of Polynesian peoples from previously settled areas of the Austronesian-speaking world, and none in support of an origin elsewhere (Howe, 2003).

The first major population movement away from the Austronesian homeland was that which took speakers of Proto-Malayo-Polynesian out of the Taiwan area into the Philippines, presumably via an entry point in the north. A series of population movements and associated linguistic splits would have seen this original group spread to the rest of that archipelago and ultimately to all of what is now Indonesia. As part of this series of population movements, the island of Madagascar was settled by a group of people who originated from Kalimantan.

Ultimately, speakers of a language immediately ancestral to Proto-Oceanic moved out of the area of Halmahera and the Indonesian province of Papua into the Oceanic homeland of New Britain and New Ireland approximately 3500 years ago. Melanesia was at the time occupied by speakers of non-Austronesian languages. As early descendants of Proto-Oceanic speakers began to spread, their progress was limited on the mainland of New Guinea to coastal areas, while the hinterland and the distant interior continued to be occupied by speakers of non-Austronesian languages.

However, there was obviously substantial linguistic contact between these two major linguistic groups, as the reconstructible VO word order often shifted to OV order in this area under the apparent influence of non-Austronesian patterns. A small number of languages – most famously Magori and Maisin of Papua New Guinea – underwent such thoroughgoing influence from non-Austronesian languages that for many years there was dispute as to whether they should be classified as Austronesian or not.

Population movements – and associated linguistic diversification – continued with the eastward drift of Oceanic-speaking groups. Up to this stage, most of the population movements involved relatively short ocean voyages to islands that were clearly visible from neighboring populated islands. However, about 3000 years ago there was a period of sudden geographical expansion out of western Melanesia involving a series of major ocean voyages into what must initially have been unknown and unpopulated territory. These voyages led to settlement as far afield as Tonga and Samoa.

It was in Tonga and Samoa that Proto-Polynesian diverged from its ancestor. From there, an even more dramatic series of oceangoing voyages led to the discovery and settlement of nearly every island group in the Pacific between the time of the birth of Christ and A.D. 1000. It should be kept in mind that this was all happening at a time when Britons were still lumbering across narrow rivers in coracles. It was Polynesian sailors who were the world's first major navigators, rather than the likes of Vasco da Gama, Christopher Columbus, and Captain Cook, whose voyages followed well over 1000 years later.

Possible External Genetic Relationships

The Austronesian languages constitute a very well defined language family in that there are few languages whose status as being Austronesian is in dispute. The status of some languages has been the subject of debate given the possibility of influence from so-called Papuan or non-Austronesian languages in the

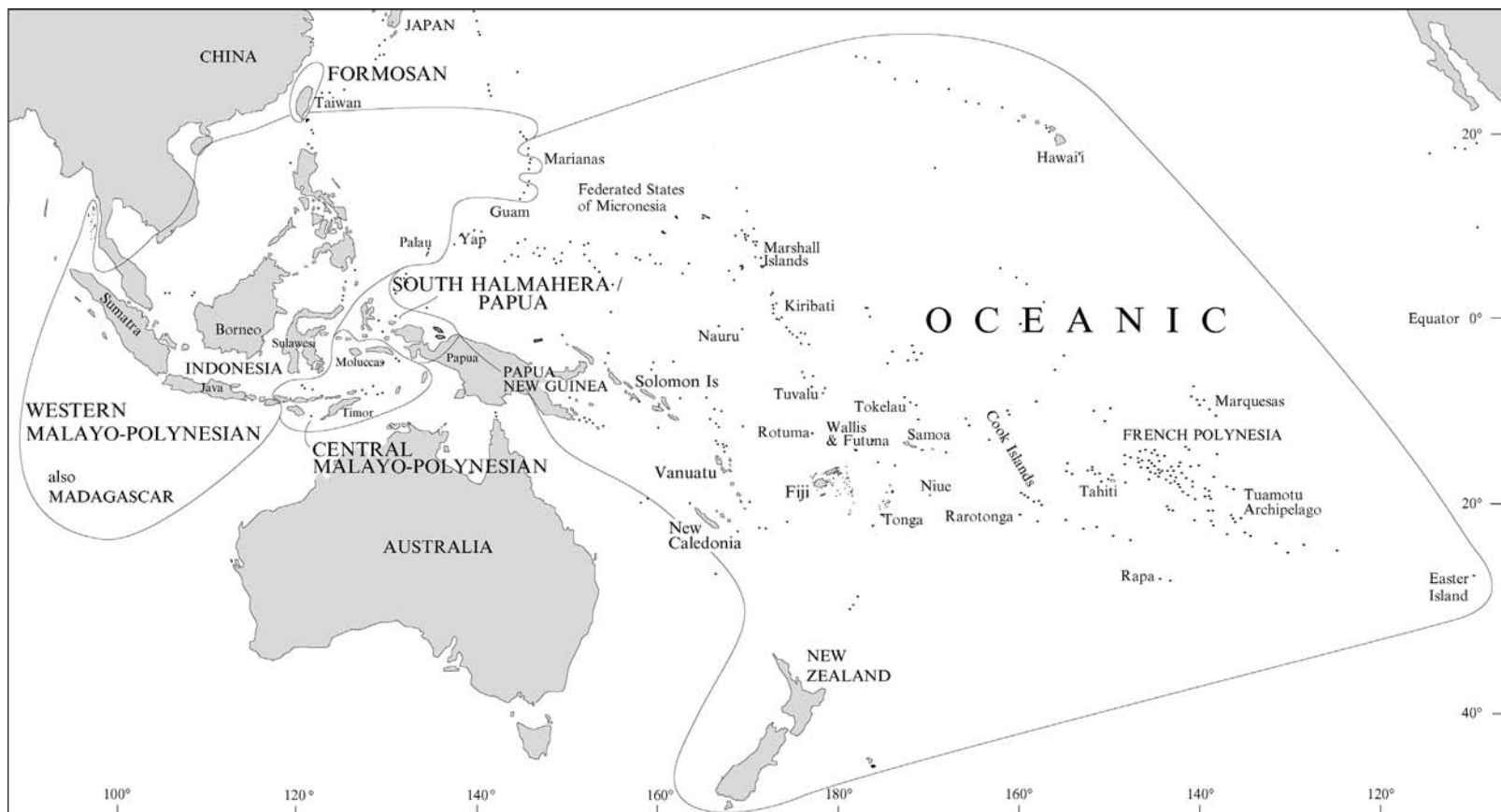


Figure 1 The Austronesian family and major Austronesian language groups (drawn by Malcolm Ross).

New Guinea area, and different scholars in the past were uncertain as to whether Maisin and Magori should be treated as non-Austronesian languages or as Oceanic languages (Lynch *et al.*, 2002: 16). While both clearly show evidence of substantial borrowing from non-Austronesian languages, it is now accepted that they can be treated as genuinely Austronesian languages. There are also a handful of languages spoken in the extreme east of Solomon Islands about which there has been some debate in the past; this issue has not yet been definitively resolved.

Most suggestions of relationships between Austronesian languages and other language groupings call for greater willingness to come to conclusions on the basis of relatively little evidence. Some scholars have claimed that there is a relationship between Austronesian languages and Japanese, others between Austronesian and the Tai-Kadai languages of southern China, others between Austronesian and Sinitic languages, and others between Austronesian and Austro-Asiatic languages, such as Nicobarese (Ross, 1994: 95–99). The few similarities between Japanese and Austronesian seem likely, at best, to involve a few very early Austronesian loans into Japanese. The other links may point to a series of relationships within a single very large grouping, though few would regard such a hypothesis as demonstrable.

Just as interesting, of course, is the question of a possible relationship between Austronesian languages and language groupings for which no suggestions of wider relationships have ever been offered. Most significant among these are the Australian languages, which represent a completely separate language family in their own right, and the various ‘Papuan’ languages of the New Guinea area. So convincing is the lack of relationship between the latter and Austronesian languages that these ‘Papuan’ languages are often collectively referred to in regional studies simply as non-Austronesian languages.

Bibliography

- Bellwood P, Fox J J & Tryon D (eds.) (1995). *The Austronesians: historical and comparative perspectives*. Canberra: Department of Anthropology (Research School of Pacific and Asian Studies, Australian National University).
- Blust R A (1970). ‘Proto-Austronesian addenda.’ *Oceanic Linguistics* 8, 85–104.
- Blust R A (1980). ‘Austronesian etymologies.’ *Oceanic Linguistics* 19, 1–181.
- Blust R A (1999). ‘Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics.’ In Zeitoun E & Li P J-K (eds.) *Selected papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academic Sinica. 31–94.
- Crystal D (1987). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Dempwolff O (1934). *Vergleichende Lautlehre des Austronesischen Wortschatzes, 1, Induktiver Aufbau einer Indonesischen Ursprache. Beihefte zur Zeitschrift für Eingeborenen-Sprachen* 15. Berlin: Dietrich Reimer.
- Dempwolff O (1937). *Vergleichende Lautlehre des Austronesischen Wortschatzes, 2: Deduktive Abwendung des Urindonesischen auf Austronesische Einzelsprachen. Beihefte zur Zeitschrift für Eingeborenen-Sprachen* 17. Berlin: Dietrich Reimer.
- Dempwolff O (1938). *Vergleichende Lautlehre des Austronesischen Wortschatzes, 3: Austronesisches Wörterverzeichnis. Beihefte zur Zeitschrift für Eingeborenen-Sprachen* 19. Berlin: Dietrich Reimer.
- Dyen I (1951). ‘Proto-Malayo-Polynesian *Z.’ *Language* 27, 534–540.
- Dyen I (1953a). ‘Dempwolff’s *R.’ *Language* 29, 359–366.
- Dyen I (1953b). *The Proto-Malayo-Polynesian laryngeals*. Baltimore: Linguistic Society of America.
- Dyen I (1965). ‘Formosan evidence for some new Proto-Austronesian phonemes.’ *Lingua* 14, 285–305.
- Grace G (1969). ‘A Proto-Oceanic finderlist.’ *Working Papers in Linguistics, University of Hawai‘i* [1/] 2, 39–84.
- Grimes B F, Grimes J E, Ross M D, Grimes C E & Tryon D T (1994). ‘Listing of Austronesian languages.’ In Tryon D T (ed.) *Comparative Austronesian dictionary: an introduction to Austronesian studies*. Berlin: Mouton de Gruyter. 121–279.
- Howe K R (2003). *The quest for origins: who first discovered and settled in New Zealand and the Pacific Islands?* Auckland: Penguin Books.
- Lynch J & Crowley T (2001). *Languages of Vanuatu: a new survey and bibliography*. Canberra: Pacific Linguistics.
- Lynch J, Ross M & Crowley T (2002). *The Oceanic languages*. Richmond, UK: Curzon Press.
- Milne W (1968). ‘Proto-Oceanic addenda.’ *Oceanic Linguistics* 7/2, 147–171.
- Pawley A K & Ross M (1995). ‘Austronesian historical linguistics and culture history.’ *Annual Review of Anthropology* 22, 425–459.
- Reid L A (1978). ‘Problems in the reconstruction of Proto-Philippine construction markers.’ In Wurm S A & Carrington L (eds.) *Second International Conference on Austronesian Linguistics: Proceedings*. Canberra: Pacific Linguistics. 33–66.
- Ross M (1988). *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.
- Ross M (1994). ‘Some current issues in Austronesian linguistics.’ In Tryon D T (ed.) *Comparative Austronesian dictionary: an introduction to Austronesian studies*. Berlin: Mouton de Gruyter. 45–120.
- Ross M, Pawley A & Osmond M (1998). *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society, vol. 1: Material culture*. Canberra: Pacific Linguistics.
- Ross M, Pawley A & Osmond M (2003). *The lexicon of Proto Oceanic: the culture and environment of ancestral*

Oceanic society, vol. 2: The physical environment. Canberra: Pacific Linguistics.

Starosta S (1985). 'Verbal inflection versus deverbal nominalization in PAN: the evidence from Tsou.' In Pawley A & Carrington L (eds.) *Austronesian linguistics at the 15th Pacific Science Congress*. Canberra: Pacific Linguistics. 281–312.

Walsh D & Biggs B (1966). *Proto-Polynesian word list 1*. Te Reo Monographs. Auckland: Linguistic Society of New Zealand.

Wolff J U (2003). 'The sounds of Proto Austronesian.' In Lynch J (ed.) *Issues in Austronesian historical phonology*. Canberra: Pacific Linguistics. 1–14.

Austro-Tai Hypotheses

L A Reid, University of Hawai'i, Honolulu, HI, USA

© 2006 Elsevier Ltd. All rights reserved.

Austro-Tai is the name given to the hypothesis that the Austronesian language family and the Tai-Kadai language family are genetically related. Austronesian languages are primarily spoken in Taiwan, island Southeast Asia and the Pacific, while Tai-Kadai languages are spoken in mainland Southeast Asia, specifically South China, Vietnam, Laos, Thailand, Burma, and Assam (India). Siamese, or Thai, the national language of Thailand, is the best-known Tai-Kadai language and the largest in terms of numbers of native speakers. Lexical similarities between Thai and the Austronesian languages have long been recognized (Schlegel, 1901). However, the hypothesis that the language family to which Thai belongs is genetically related to Austronesian was first proposed by Benedict (1942). Benedict proposed that the genetic relationship between the two language families was a sister relationship, implying that the Tai-Kadai languages are the descendant languages of a parent language – Proto-Austro-Tai – from which pre-Austronesian peoples split in a move to Taiwan, where Proto-Austronesian developed.

Although his later work (Benedict, 1975) was well received by archaeologists and prehistorians, it was generally less well received by linguists, who were skeptical of the extensive array of Proto-Austro-Tai reconstructions that he proposed and his unorthodox methodology for reconstructing them. A number of critical reviews of his work appeared (esp. Gedney, 1976), casting doubt on the nature of the relationship, and pointing to a number of unrecognized loans from Chinese. But the presence in the Tai-Kadai family of a considerable number of forms from the area of basic vocabulary that are very similar in sound and meaning to corresponding Austronesian forms removes the possibility of coincidence as a possible explanation for the similarities. Whether the similarities reflect a genetic relationship or are the result

of contact was examined by Thurgood (1994). He concluded that the sound correspondences and tonal developments within the Tai-Kadai languages of forms with comparable Austronesian reconstructions are irregular and thus cannot be evidence of a genetic relationship, but rather of an early contact relationship.

A recent reexamination (Ostapirat, 2000) of the whole question of the internal relationships within the Tai-Kadai family of languages (renamed by Ostapirat as Kra-Dai) has opened up once again the nature of the relationship that exists between this family and Austronesian. In an insightful paper, Ostapirat (2005) presented a list of some 50 pan-Kra-Dai basic vocabulary items, at least half of which can be related by regular sound correspondences to equivalent forms in Proto-Austronesian. The English glosses of Kra-Dai forms that appear to be related to Austronesian include the following: *bear* (n.), *bird*, *black*, *child*, *eat*, *excrement*, *eye*, *far*, *fire*, *grandmother*, *grease*, *hand*, *head*, *I*, *leaf*, *leg*, *live*, *louse* (head), *moon*, *nose*, *raw*, *sesame*, *shoulder*, *this*, *tooth*, *water*, and *you*.

From this evidence, Ostapirat concluded, "It does not seem likely that the very high number of roots between Kra-Dai and Austronesian that emerge from the core list could be accidental or simply result from borrowings." In commenting on Thurgood's claims that there are no regular sound correspondences between the Kra-Dai and Austronesian families, Ostapirat explained that they are the result of Thurgood's being unaware of crucial data from little-known languages, and of the inadequacy of some of his Proto-Kra-Dai reconstructions. Despite the apparent strength of the evidence he cited, Ostapirat nevertheless considered the evidence to be debatable as proof of a genetic relationship between the families.

An alternate hypothesis regarding the external relationships of the Kra-Dai family is that the languages are genetically related not to Austronesian, but to the Sino-Tibetan family. Ostapirat rejected this hypothesis, noting that etyma that appear to be related to Chinese are rarely found in all branches of the family

and almost none belong to the core vocabulary of the language.

The most recent view of the external relationships of the Kra-Dai family is that proposed by Sagart (2005). Building on the comparisons established by Ostapirat, Sagart presented data from a recently described language of the Kra group, Buyang (BYU) (Li, 1999), which apparently, alone, among the Kra-Dai family, retains a number of disyllabic forms which correspond to Proto-Austronesian or Proto-Malayo-Polynesian (PMP) reconstructions, such as BYU *ma⁰ tɛ⁵⁴* 'die' (PMP *matay); BYU *ma⁰ ta⁵⁴* 'eye' (PMP *mata); BYU *qa⁰ ðu¹¹* 'head' (PMP *quluh); BYU *ma⁰ ðu³¹²* 'eight' (PMP *walu), etc.

Data such as these establish beyond any doubt that a genetic relationship exists between the two families. The nature of the relationship, however, is still being discussed. Sagart rejected the possibility that Proto-Kra-Dai and Proto-Austronesian are sister languages, thereby rejecting the Austro-Tai hypothesis in its original formulation. He claimed instead that the Kra-Dai languages are a subgroup of Austronesian, being descendants of the language spoken by a group of Austronesian-speaking people who returned to the mainland from the east coast of Taiwan, long after the first Austronesian settlement there, but probably before the movement south to the Philippines some 4000 years ago of the ancestors of the Extra-Formosan, or Malayo-Polynesian, languages.

Ostapirat stated that if Kra-Dai were a subgroup within Austronesian, as Sagart believes, it would seem likely that they must have belonged to one of the primary branches, in that Proto-Kra-Dai retains a distinction between the reflexes of Proto-Austronesian *C and *t, and *N and *n, pairs of sounds which fell together and are therefore not distinguished in Proto-Extra-Formosan. Moreover Proto-Kra-Dai retains a sibilant reflex of Proto-Austronesian *S, which developed as *h in Proto-Extra-Formosan. In addition, he noted that although there are no Extra-Formosan languages which have reflexes of Proto-Austronesian *Cumay 'bear (n.),' the form is reflected in Kra-Dai languages.

Sagart's position, on the other hand, is that the ancestors of the Kra-Dai languages must have

returned to the mainland about the time that the ancestors of the Extra-Formosan languages moved south, in that they apparently reflect certain forms, such as *-mu 'you (sg.),' *lima 'five,' *manuk 'chicken,' etc., that Sagart believes are reconstructible only to the parent of the Extra-Formosan languages, but not to Proto-Austronesian.

If the speakers of the parent of Proto-Tai-Kadai did in fact return to the mainland from Taiwan as proposed by Sagart, he suggests that they probably settled in coastal areas in Guangdong or Guangxi, and their language was eventually relexified by a language from some probably extinct phylum, but one ultimately related to Austroasiatic, retaining only the most basic elements of its Austronesian lexicon.

Bibliography

- Benedict P K (1942). 'Thai, Kadai, and Indonesian: a new alignment in southeastern Asia.' *American Anthropologist* 44, 576–601.
- Benedict P K (1975). *Austro-Thai: language and culture, with a glossary of roots*. New Haven: HRAF Press.
- Gedney W J (1976). 'On the Thai evidence for Austro-Tai.' *Computational Analyses of Asian and African Languages* 6, 65–82.
- Li J (1999). *Buyang yu yanjiu*. Beijing: Zhongyang Minzu Daxue.
- Ostapirat W (2000). 'Proto-Kra.' *Linguistics of the Tibeto-Burman Area* 23(1).
- Ostapirat W (2005). 'Kra-dai and Austronesian: notes on phonological correspondences and vocabulary distribution.' In Sagart *et al.* (eds.). 107–131.
- Sagart L (2005). 'Tai-Kadai as a subgroup of Austronesian.' In Sagart *et al.* (eds.). 177–181.
- Sagart L, Blench R & Sanchez-Mazas A (eds.) (2005). *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London/New York: Routledge-Curzon.
- Schlegel G (1901). 'Review of Frankfurter's Siamese grammar.' *T'oung Pao* 2, 76–87.
- Thurgood G (1994). 'Tai-Kadai and Austronesian: the nature of the historical relationship.' *Oceanic Linguistics* 33(2), 345–368.

Avestan

E Tucker, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Avestan is the language of the most ancient collection of texts sacred to the Zoroastrian religion. It represents the Old Iranian stage of the Iranian language family, and provides, along with Old Persian, the earliest evidence for the Iranian branch of the Indo-Iranian and Indo-European language families. The language is known only via a defective medieval manuscript tradition, which was preserved in Iran and also in India, to which adherents of the Zoroastrian faith, the Parsis, emigrated according to tradition in 10th-century A.D. Hence its study presents a range of philological, textual, and interpretative problems.

Two forms of the language are documented: Old Avestan (OAv.; sometimes called Gathic Avestan) and Younger Avestan (YAv.). Opinion is at present divided as to whether they represent earlier and later forms of precisely the same language, or whether dialect differences are also involved. A difficulty is that OAv. is known only from a very small number of texts: the *Gāthās*, seventeen complex poems in five different meters attributed to the prophet Zarathuštra himself, which still present many problems of interpretation; two short sacred prayers; and a traditional liturgy in seven sections, the *Yasna Haptañhāiti*. The most extensive surviving Avestan texts, that is, the other parts of the 72-chapter *Yasna*, the lengthy *Yāšt*s, which honor divinities such as Mithra and Anāhitā, and the *Vīdēvdāt*, ‘the Law which rejects False Gods’ are composed in YAv. A few short sections of the *Yasna* are in YAv. with an artificial veneer of OAv. phonological features (pseudo-OAv.). Faulty grammar in parts of the *Avesta* may suggest that composition continued at a stage when Avestan was no longer a living language, but the text may also have deteriorated during transmission.

Absolute dates for Avestan are entirely lacking. The date of the prophet Zarathuštra is still debated, but many scholars agree that the *Gāthās* must be roughly contemporary with the *RigVeda* in India (i.e., toward the end of the second millennium B.C.), as OAv. morphology and syntax are on a par with those of the earliest Vedic language. YAv. shows many simplifications, particularly in its verb system, and innovations, and the text collection as a whole must span several centuries.

Avestan diverges from Old Persian in some important sound changes (IE **k̑, g̑, gh̑* > Av. *s, z, z*, but > OP *θ, d [ð], d [ð]*; IE **k̑u, g̑u* > Av. *sp, zb*, but OP *s, z*),

but in these respects it agrees with the majority of Iranian languages. It is often described as East Iranian because geographical names found in the YAv. texts refer to the region of present day Afghanistan and East Iran and none refer to West Iran. However, Avestan does not share in the most characteristic features known from Middle Iranian languages of the extreme East, such as Khotanese. Rather it shows several phonological developments (if these do indeed belong to the original Avestan language) that are unparalleled elsewhere in Iranian (**-aha-* > *-aṇha-*, **-ft-* > *-pt-*, **-rt-* > *-ṣ-* when the syllable was accented, etc.). No Iranian language known from later times can be identified as a direct descendant of Avestan.

The Avestan texts were composed orally, and they were recited and transmitted orally by the Zoroastrian priesthood in different regions of Iran, but it is hard, if not impossible, to assign specific features of Avestan to influence from specific local languages. The written recension was only made during the Sasanian period (224–651 A.D.), when Zoroastrianism flourished as the state religion. An elaborate alphabet of 53 signs, including 16 for vowels, was invented on the basis of the cursive Zoroastrian Pahlavi script and the Christian Psalter script (both derived from Aramaic) in order to record as precisely as possible the traditional pronunciation of Avestan, which had ceased to be a living language several centuries earlier. Avestan orthography is not based on phonemic principles, but it conveys a wealth of information about allophonic variation. Consequently, Avestan words often look very different from their exact counterparts in Vedic, even though the languages are closely related; contrast YAv. *hāuuōiia-*: Vedic *śavya-* ‘left’ or OAv. *mānghī*: Vedic *mānsi* ‘I thought.’ Moreover, morphological regularities within the Avestan language itself are often obscured (e.g., *barahi*, *baraiti*, *barenti* represent 2 sg., 3sg., 3pl. present active based on the inherited thematic stem *bara-* ‘bear,’ cf. Skt. *bharasi*, *bharati*, *bharanti*).

The Avestan manuscripts, of which the earliest dates from 13th-century A.D., reflect a written tradition that barely survived the centuries following the Islamic conquest. At one stage only a single manuscript existed for each part of the extant *Avesta*, and approximately three-quarters of the *Avesta* as described in the Sasanian Zoroastrian books has been lost. Recent scholarship has made progress in reconstructing the spellings of the ‘Sasanian Archetype’ text, but it is still often difficult to determine which features belong to the original Avestan language and which arose in the course of either oral or

written transmission. According to some scholars, the phonology of OAv. was close to that of Proto-Iranian, and Gathic meter may provide evidence for archaic features such as the vocalization of semivowels according to Sievers' Law, and a hiatus between vowels caused by the recent loss of laryngeals.

The inflectional morphology of both OAv. and YAv. is extremely rich. For nouns, adjectives and pronouns, the full set of eight IE case inflections and three numbers remain alive, with a huge range of nominal stem types, and some ancient irregular paradigms, such as YAv. *paṇtā* (nom.), *paṇō* (gen.) 'path'; OAv. *huuarē* (nom.), *x^uəng* (gen.) 'sun'; OAv. *aogō* (nom.), *aojaṇhā* (instr.) 'strength.' The OAv. enclitic acc. pl. personal pronouns *nā* 'us,' *vā* 'you' (cf. Latin *nōs*, *vōs*) are an archaism not found elsewhere in Indo-Iranian. At the same time, there are innovations, such as the OAv. (and YAv.) nom. pl. masc. ending *-ā* in thematic stems (more frequent than inherited *-ā*, *-āṇhō*), and the creation in YAv. of a distinct ablative singular inflectional ending for all nominal classes.

In the OAv. verb system all the IE tense-aspect stems (present, aorist, perfect) are fully employed. YAv. has a much simplified system where present and preterite are based on a single stem (the inherited present) and distinguished by different inflectional endings. The inherited augment *a-* rarely appears, and its function in Av. is problematic. Although thematic presents are productive, the rarer types of athematic present are well represented, notably acrostatic root presents

(OAv. *stāumī* 'I praise,' *aogədā* 'he said,' YAv. *āṇhāire* 'they sit'). Modal forms (subjunctives, optatives, and imperatives) are frequent at all stages.

The Avestan lexicon is remarkably free of loan-words from non-Iranian languages, and it preserves some IE lexemes that were lost in Indo-Aryan, e.g., *varəz-* 'to work,' *vad-* 'to lead.' Contrasting vocabulary items for good (ahuric) versus evil (daevic) beings reflect Zoroastrian dualism but their linguistic origins are complex (e.g., *staman-/zafar-* 'mouth,' *dōiṇra-/aši-* 'eye,' *aog-/dauu-* 'to speak,' *tak-/zbar-* 'to run,' *nmāna-/gərəda-* 'house,' *ṇṇərəs-/karət-* 'to fashion').

Bibliography

- Bartholomae C (1904). *Altiranisches Wörterbuch*. Strassburg: Trübner.
- Hoffmann K (1987). 'Avestan language.' In Yarshater E (ed.) *Encyclopedia Iranica*, vol. III/1. London: Routledge. 47–62.
- Hoffmann K & Narten J (1989). *Der Sasanidische Archetypus: Untersuchungen zu Schreibung und Lautgestalt des Avestischen*. Wiesbaden: Reichert.
- Hoffmann K & Forssmann B (1996). *Avestische Laut- und Flexionslehre*. Innsbruck: Institut für Sprachwissenschaft.
- Kellens J (1989). 'Avestique.' In Schmitt R (ed.) *Compendium linguarum Iranicarum*. Wiesbaden: Reichert. 32–55.
- Sims-Williams N (1998). 'The Iranian Languages.' In Giacalone Ramat A & Ramat P (eds.) *The Indo-European languages*. London: Routledge. 125–163.

Aymará

W F H Adelaar, Leiden University, Leiden,
The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The name 'Aymara' is used for one of the most important native languages of South America. It is spoken by approximately 2 000 000 people in three countries: Bolivia (mainly in the department (administrative division) of La Paz, but also in parts of Cochabamba, Oruro, and Potosí), Chile (in the highlands of Tarapacá), and Peru (in the departments of Moquegua, Puno, and Tacna). Aymara is closely related to the Jaqaru language – spoken by less than a 1000 people, mainly in the village of Tupe in the province of Yauyos (department of Lima) in central Peru – as well as to Cauqui, spoken by a

few individuals in the nearby village of Cachuy. Together, the languages Aymara, Jaqaru, and Cauqui form a family that has variously been called 'Jaqi' (Hardman, 1978), 'Aru' (Torero, 1972), 'Aimara' (Cerrón-Palomino, 2000), and 'Aymaran,' which is the name used in this article.

The Aymaran language family has no proved external relatives. There are close and detailed similarities in the phonological, structural, and lexical domains with the neighboring Quechua language group; the two groups also share more than 20% of their lexicon. This situation suggests a protracted period of interaction between the underlying protolanguages of both Aymaran and Quechua. The interaction may have continued on a local basis during the further development and expansion of both language groups. The close similarities between the two language

groups have often been interpreted as a proof of common origin (the so-called Quechumaran hypothesis). Nevertheless, most similarities are attributable to linguistic convergence, making it difficult to distinguish between borrowed and inherited material (see *Andean Languages; Quechua*).

Before 1600, Aymara and related dialects were widely spoken in southern Peru and in the eastern and southern Bolivian highlands, where Quechua is now the dominant language. The historical influence of Aymara through borrowing can be appreciated from the spread of Aymara numerals into the southern cone of South America (Mapuche) and into the Amazonian basin (Tacanan languages). The name 'Aymara' is probably derived from a province or ethnic group located in the present-day Peruvian department of Apurímac (now Quechua speaking). The study of the Aymara language received an important stimulus in the 17th century when the Jesuit order established a mission in Juli, on the southwestern shore of Lake Titicaca. The first grammar and dictionary of Aymara were written in 1603 and 1612, respectively, by a Jesuit, Ludovico Bertonio.

The Aymara vowel system consists of three vowels (*a, i, u*), of which the high vowels are lowered to (*e, o*) next to a uvular consonant. There is a distinction of vowel length that is mainly used in morphology, but also in a few lexical roots. Stops and affricates are normally voiceless; they can be plain, glottalized, or aspirated. There is a contrast between velar and uvular consonants. Dialects in the border area of Bolivia, Chile, and Peru have a distinctive velar nasal consonant. Stress is predictable and is located on the penultimate syllable or mora. All roots are vowel final. However, the final vowel of a nominal expression is regularly deleted before pause. Although the structure of Aymara roots and suffixes is basically simple, surface forms can be complex due to the fact that many suffixes trigger the suppression of a preceding vowel. This suppression must be treated as a formal property of the suffix in question, because there are no synchronically valid phonological rules to account for it. In some cases, root-interior vowels are also suppressed under similar circumstances. These different types of vowel suppression produce elaborate consonant clusters, as illustrated in *han unxtkiri* 'without moving,' 'immobile,' which can be analyzed as follows (vowels between parentheses are suppressed; PROG, progressive; AGT, agent; NOM, nominalizer):

- (1) han(i) un(u)q(i)-t(a)-k(a)-iri
not rock-upward/begin-PROG-AGT.NOM

The combination *unxta-* (< unuqi-ta-) is fixed and is interpreted as 'to move slightly.'

Aymara has an agglutinating structure mainly based on suffixation; there are no prefixes at all. The morphology is complex, but regular. Words containing as many as nine consecutive suffixes are no exception (the fixed combination *aru-si-* means 'to speak'; INCL, inclusive; REFL, reflexive; PL, plural; COMPL, completive; BEN, benefactive):

- (2) hiwas-kam(a) aru-s(i)-kipa-si-p-ɣa-ña-naka-
taki-sa
we(INCL)- speak-REFL-turn-REFL-PL-
case:among COMPL-NOM-PL-case:BEN-too
'so that we are able to communicate among
ourselves'

Verb-final order is obligatory in dependent clauses and is the preferred order in full sentences. In noun phrases, all modifiers precede their heads. Nouns can be marked for case, number (plural), and person of possessor. The overall structure of the language is nominative-accusative. Case is expressed by suffixes, but the accusative is marked by eliminating a stem-final vowel. There is a four-term pronominal system consisting of speaker (*naya* 'I'), addressee (*huma* 'you'), third person (*hupa* 'he/she'), and an inclusive plural that comprises both speaker and addressee (*hiwasa*). This system is also reflected in nominal possession and in verbal inflection.

Verbs in Aymara exhibit a rich derivational morphology, including causative, reflexive-reciprocal, spatial direction, number of subject, aspect, speaker orientation ('hither'), and several other options. Tense, mood, and personal reference, both of the subject and of a human (in)direct object, are combined in complex portmanteau endings, which are a hurdle for the nonnative learner. In these endings (nine for each tense or mood paradigm), a third-person object is not explicitly indicated. Characteristic for the Aymara verb is the existence of evidential distinctions (inference, conjecture, nonpersonal witness, etc.), for which the Aymara society is highly sensitive.

Verbalizations – copula 'to be,' locative verb 'to be at' – are indicated morphologically, the former by vowel lengthening (1POSS, first-person possessor; VERBAL, verbalizer; 2SUB, second-person subject; ASSERT, assertive):

- (3) hič^ha-ɣ(a) wawa-ha-:-ɣ(a)-ta-wa.
now-topic child-1POSS-VERBAL :be-COMPL –
2SUB-ASSERT
'Now you are already my child.'

Nominalization plays an important central role in Aymara morphosyntax. Different types of dependent clauses are obtained by combining nominalized verbs with specific case markers. Nominalization is also

used to form relative clauses. In contrast to Quechua, person of object cannot be indicated morphologically in nominalized verbs. (Note: Examples (1)–(3) are from Albó and Layme (1992)).

Bibliography

- Adelaar W & Muysken P (2004). *The languages of the Andes*. Cambridge: Cambridge University Press.
- Albó X & Layme F (1992). *Literatura aymara: antología I: prosa*. La Paz: CIPCA, Hisbol, JAYMA.
- Bertonio L (1603). *Arte y gramática muy copiosa de la lengua aymara*. [Reprinted in 1879]. Leipzig: Teubner.
- Bertonio L (1612). *Vocabulario de la lengua aymara*. [Reprinted in 1984. Cochabamba: CERES, IFEA, MNEF.]

- Cerrón-Palomino R (2000). *Lingüística aimara*. Cuzco: Centro Bartolomé de Las Casas.
- Hardman M (1978). 'Jaqi: the linguistic family.' *International Journal of American Linguistics* 44(2), 146–153.
- Hardman M (ed.) (1981). *The Aymara language in its social and cultural context*. Gainesville: University of Florida Press.
- Hardman M (2000). *Jaquaru*. Munich: Lincom Europa.
- Hardman M, Vásquez J & de DiosYapita J (1974). *Outline of Aymara phonological and grammatical structure*. Gainesville: University of Florida Press.
- Torero A (1972). 'Lingüística e historia de la sociedad andina.' In Escobar A (ed.) *El reto del multilingüismo en el Perú*. Lima: Instituto de Estudios Peruanos. 47–106.

Azerbaijani

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Azerbaijani (Azerbaijani, Azeri) (*Azərbaycan dili*, *Azərbaycanca*) belongs, like Turkish, to the western group of the southwestern, or Oghuz, branch of the Turkic language family. It is spoken in northern and southern Azerbaijan (i.e., in the Republic of Azerbaijan), particularly in the province of Azerbaijan, and in Iran. Azerbaijani is the official language of the Republic of Azerbaijan (*Azərbaycan Respublikası*), which constitutes the easternmost part of Transcaucasia. The Republic is situated between Iran and Russia, with a small European portion north of the Caucasus range. It includes the exclave of the Nakhchivan Autonomous Republic and the separatist Nagorno-Karabakh region. It borders on the Russian Federation in the north, Georgia in the northwest, Armenia in the west, Iran in the south, and the Caspian Sea in the east. Azerbaijanians make up about 90% of the Republic's total population of about 7.8 million. Other ethnic groups include Dagestanis, Russians, and Armenians (mainly in Nagorno-Karabakh). Over 80% of the citizens speak Azerbaijani as their first language. The number of speakers in the Republic amounts to about 7 million. The standard language is based on the dialect of the capital Baku (*Bakı*). The number of speakers in southern Azerbaijan, which is located in northwestern Iran and borders on Turkey in the west,

is estimated to be over 13 million. Similar varieties are spoken in eastern Anatolia, northern Iraq, Georgia, and Armenia. The total number of speakers may amount to 20 million.

The current status of the language in the Republic is very solid. More than half of Azerbaijani speakers are monolingual. The social situation of the varieties of Azerbaijani spoken in Iran is quite different. There the languages have not been promoted; on the contrary, their use has been discouraged and public use of Azerbaijani was banned for several decades. The situation is now improving.

Origin and History

The language goes back to the Oghuz Turkic varieties of the Seljuks, who immigrated to the area in the 10th and 11th centuries. These people originally belonged to the Oghuz confederation of tribes, whose Inner Asian steppe empire collapsed in 744. Due to political and religious differences, Azerbaijanian Turks for centuries lived in relative separation from the Turks of Turkey. Azerbaijan's history shows substantial cultural influence from Iran. In 1828, Azerbaijan was divided into a northern and a southern part under Russian and Persian rule, respectively. Northern Azerbaijan was part of the former Soviet Union for 70 years. It regained independence in 1991.

Related Languages and Language Contacts

The language is related to Turkish, Gagauz, South Oghuz, Khorasan Turkic, and Turkmen. It has a

strong Iranian substrate and has for many centuries been in close direct contact with Persian. Turkish had a considerable influence on the northern Azerbaijani standard language as established before the Soviet era. During the past century, Russian has influenced the standard language, whereas the contacts with Turkish have been very limited. There is nevertheless a high degree of interintelligibility with Anatolian Turkish.

An Azerbaijani koiné functioned for centuries as a lingua franca, serving trade and intergroup communication all over Persia, in the Caucasus region and in southeastern Dagestan. Its transregional validity continued at least until the 18th century. Later on, it lost its importance in favor of Persian in the south, whereas Russian was dominant in the north. In the period of Russian domination of economy and politics, Russian had a strong position; 38% of the Azerbaijanis of the Republic still speak Russian fluently.

The Written Language

The early history of Azerbaijani as a literary language is closely linked to that of Anatolian Turkish. Signs of its detachment are found in sources written at the end of the 14th century. Azerbaijani has a long and rich literary tradition. The language was written in Arabic script up to the 20th century. In 1923, a Latin-based script, *yaqalif* ‘the new alphabet,’ was introduced in Soviet Azerbaijan. It was a model for the Roman alphabet that was introduced in Turkey in 1928. This alphabet was replaced by a Cyrillic script in 1939–1940. In 1991, after the disintegration of the Soviet Union, the Republic of Azerbaijan adopted a new modified Roman-based alphabet incorporating a few special letters. The transition to this script has been gradual. The Republic still applies a dual script system, with the Roman- and Cyrillic-based letters appearing side by side. In southern Azerbaijan, where the written use of the language is highly restricted, the Arabic script is still used.

Distinctive Features

The language exhibits most linguistic features typical of the Turkic family (see *Turkic Languages*). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following discussions, only a few distinctive features will be dealt with – in particular, some ways in which Azerbaijani is different from Turkish. In the notation of suffixes, capital letters indicate phonetic variation, e.g., A = a/e, I = i/i. Segments in

parentheses occur after vowel-final or consonant-final stems. Hyphens are used here to indicate morpheme boundaries.

Phonology

Unlike Turkish, Azerbaijani has a mid vowel phoneme *e* and a higher phoneme *é* (e.g., *él* ‘people, country’ vs. *el* ‘hand’ and *én* ‘width’ vs. *en* ‘most’). In words of Arabic–Persian origin, non-high-position vowels are more fronted than they are in Turkish (e.g., *teref* ‘side’ vs. Turkish *taraf*). Common Turkic initial *y-* is often lost before high vowels (*üz* ‘face’ (Turkish *yüz*) and *ulduz* ‘star’ (Turkish *yıldız*)). Initial *ï-* is replaced by *i-* (*il* ‘year’ (Turkish *yıl*)). Vowels are often rounded in the neighborhood of *υ* (*ov* ‘hunt’ (Turkish *av*)).

The spoken language is relatively conservative with respect to sound harmony. It still displays invariable suffixes – i.e., suffixes not subject to sound harmony (*gel-dox* [come-PAST-1.PL], *gel-dix* [come-PAST-1.PL] ‘we came’ and *işle-max* [work-INF] ‘to work’); cf. Turkish *gel-dik* [come-PAST-1.PL] and *işle-mek* [work-INF] (with front–back and rounded–unrounded harmony). In the standard language, the vowel harmony is normalized on the standard Turkish model, e.g., *it-ler-imiz-den* [dog-PL-POSS.1. PL-ABL] (front vowels) ‘from our dogs’ vs. *at-lar-ımız-dan* [horse-PL-POSS.1.PL-ABL] (back vowels) ‘from our horses.’ A few suffixes are invariable. As in Turkish, rounded vs. unrounded harmony does not affect low suffix vowels.

Common Turkic initial *q-* is, as in Turkmen, represented by the back-voiced stop *g*-, e.g., *ğara* ‘black’ (Turkish *kara*). Common Turkic final back *-q* is represented by *-g* in polysyllabic words and in certain monosyllabic words (after originally long vowels), e.g., *ayag* ‘foot’ (Turkish *ayak*), *ag* ‘white’ (Turkish *ak*). It is fricativized to *-x* in other cases (*yox* ‘non-existent’ (Turkish *yok*)). Stem-internal *q* is also fricativized (*yaxın* ‘near’ (Turkish *yakın*)). The voicing of Common Turkic *k-* generally follows the same pattern as in Turkish (*gör-* ‘to see’ < *kör-*). There are, however, some differences, as for *kêç-* ‘to pass’ vs. Turkish *geç-*. The distribution of the initial dentals *t-* and *d-* is generally the same as in Turkish (*dış* ‘tooth’ < *ti:ş*). Exceptions include *tik-* ‘to sew’ (Turkish *dik-*) and *daş* ‘stone’ (Turkish *taş*). The distribution of the initial labials *p-* and *b-* mostly follows the Turkish pattern. Exceptions include *barmag* ‘finger’ (Turkish *parmak*) and *poz-* ‘to destroy’ (Turkish *boz-*). As in most Turkic languages, the initial nasal *m-* occurs instead of *b-* as a result of assimilation to a following nasal (*min* ‘thousand’ (Turkish *bin*)). Glottal *h* and uvular *x*, which have merged into *h* in Turkish, are

distinct phonemes (e.g., *heyat* ‘life’ (Turkish *hayat*) and *xeber* ‘information’ (Turkish *haber*)). A word-medial glottal stop occurring in loans of Arabic origin may be pronounced or realized as vowel length, as in *te’sir* or *te:sir* ‘influence.’ Unvoiced obstruents may be strongly aspirated, as in *t^hop^h* ‘gun, cannon.’ The stops *k* and *g* are strongly palatalized in many dialects. Consonant metathesis is a rather common phenomenon (*ireli* ‘front’ (Turkish *ileri*) and *körpü* ‘bridge’ (Turkish *köprü*)).

Grammar

The dative forms of the pronouns *men* ‘I’ and *sen* ‘you’ are *mene* [I-DAT] and *sene* [you-DAT] (Turkish *bana, sana*). The marker *-(y)Ar*, corresponding to Turkish *-(V)r*, forms a general, less focused present tense with habitual, intentional, prospective, and similar meanings (e.g., *bil-er* [know-AOR] ‘knows, will know’ (Turkish *bil-ir* [know-AOR]) and *gel-er* [come-AOR] ‘comes, will come’ (Turkish *gel-ir* [come-AOR])). The present-tense marker *-(y)Ir* corresponds to Turkish *-Iyor* and Turkmen *-yA:r*, as in *yaz-ir* [write-PRES] ‘writes, is writing’ (Turkish *yaz-iyor* [write-PRES]) and *iste-yir* [want-PRES] ‘wants’ (Turkish *istiyor* [want-PRES]). Unlike in Turkish, low vowels have thus been generalized in *-(y)Ar*, whereas high vowels have been generalized in *-(y)Ir*. The first-person copula suffixes of the pronominal type are *-(y)Am* (e.g., *gör-ür-em* [see-PRES-1.SG] ‘I see,’ *al-ir-am* [take-PRES-1.SG] ‘I take,’ *gör-ür-ük* [see-PRES-1.PL] ‘we see,’ and *al-ir-iğ* [take-PRES-1.PL] ‘we take’ vs. Turkish *gör-üyor-um* [see-PRES-1.SG], *al-iyor-um* [take-PRES-1.SG], *gör-üyor-uz* [see-PRES-1.PL], and *al-iyor-uz* [take-PRES-1.PL]). The second-person singular copula suffix is *-sAn*, as in *gözel-sen* [beautiful-2.SG] ‘you are beautiful’ (Turkish *güzel-sin* [beautiful-2.SG]). The perfect paradigm contains first-person forms with *-mİş*, whereas *-(y)İb* is used in the second and third persons (e.g., *gel-miş-em* [come-PERF-1.SG] ‘I have come,’ *gel-ib-sen* [come-PERF-2.SG] ‘you have come,’ and *gel-ib-[dir]* [come-PERF-3.SG] ‘has come’). The perfect markers are not used as the corresponding Turkish *-mİş* markers, which have indirective meaning. Thus, forms such as *goy-muş-am* [put-PERF-1.SG] ‘I have put’ and *al-miş-am* [take-PERF-1.SG] ‘I have taken’ are translated into Turkish by *koy-d-um* [put-PAST-1.SG] and *al-d-im* [take-PAST-1.SG] rather than by *koy-muş-um* [put-EV-1.SG] and *al-miş-im* [take-EV-1.SG]. The Persian influence on the dialects varies considerably. Some varieties use the comparative suffix *-ter* and the superlative suffix *-teri:n*, both copied from Persian.

Though the syntax is rather similar to that of most other Turkic languages, the Persian impact has been

considerable, especially in the southern varieties. Many conjunctions and other functional words are copied from Persian and Arabic (via Persian), e.g., *ki*, which precedes complement and relative clauses.

Lexicon

Due to the different political and cultural developments for the past 600 years, the Azerbaijani vocabulary differs from the modern Turkish vocabulary in many respects. There are certain differences in the genuinely Turkic lexicon (*tap-* ‘to find’ vs. Turkish *bul-*, *öz* ‘self’ vs. *kendi*, *isti* ‘warm’ vs. *sıcak*, *düş* ‘to go down, to land’ vs. *in-*, *sümük* ‘bone’ vs. *kemik*). Turkish *düş-* means ‘to fall’; *sümük* means ‘mucus.’ The vocabulary has preserved numerous elements of Persian and Arabic–Persian origin that have been abandoned in Turkish as a result of the puristic language reforms, including *lüyet* ‘dictionary’ (Turkish *sözlük*), *muellim* ‘teacher’ (*öğretmen*), and *pul* ‘money’ (*para*).

Since the 19th century, Russian loanwords, particularly technical terms, have entered the northern Azerbaijani varieties (*zavod* ‘factory’ (Turkish *fabrika*), *fevral* ‘February’ (Turkish *şubat*), *stul* ‘chair’ (Turkish *sandalye*), and *galstuk* ‘necktie’ (Turkish *kra-vat*)). The southern varieties exhibit many loans from Persian (e.g., *miz* ‘table’ and *ruzname* ‘newspaper’ (the northern varieties have *stol* and *gezet*)).

Dialects

The spoken language includes several dialects. They are mostly divided into three groups: northern dialects spoken in the Republic of Azerbaijan, southern dialects in northwestern Iran, and East Anatolian dialects. Though these dialects differ a great deal from each other, they are mostly mutually intelligible. Among the northern dialects, there is a western subgroup in the central part of the Republic (including Genje, Shusha, Kazak, Karabagh, and Ayrum). Dialects of an eastern subgroup are spoken on the shore of the Caspian Sea, in Derbent, Kuba, Shemakha/Shamakhi, Baku, Sal’jany, Mughan, and Lenkoran, for example. The standard language is based on the urban dialect of the capital Baku. Dialects spoken in the northern parts of the Republic include Zakataly, Nukha, and Kutkashen. Dialects spoken in the southern parts of the republic include those of Nakhchevan and Ordubad.

The dialects of Iran include those of Tebriz, Urmia, Qūšči, Xoy, Marāya, Marand, ‘Oryān Tepe, Torkmānchay, Ardabil, Sarāb, Meyāna, and the exclave Galūgāh. The dialect of the Karapapakh ‘Black Caps’ was spoken between the upper Kura

and Arpachay Rivers, on the boundary between Armenia and Georgia, and in Persian Azerbaijan near Lake Urmiya. Some dialects are spoken in Khorasan, including Lotfābād and Daragaz.

Bibliography

- Azizbekov X A (1965). *Azerbajdžansko-russkij slovar'* Baku: Azerbajdžanskoe gosudarstvennoe izdatel'stvo.
- Doerfer G (1959). 'Das Aserbajdžanische.' In Deny J *et al.* (ed.) *Philologiae turicae fundamenta 1*. Aquis Mattiacis: Steiner. 280–307.
- Gadžieva N Z (1966). 'Azerbajdžanskij jazyk.' In Baskakov N A *et al.* (eds.) *Jazyki narodov SSSR 2. Tjurkskie jazyki*. Moskva: Nauka. 66–90.

Johanson L (2001). 'Azerbaijani.' In Garry J & Rubino C (eds.) *Facts about the world's major languages: an encyclopedia of the world's major languages, past and present*. New York, Dublin: The H. W. Wilson Company, New England Publishing Assoc. 52–54.

Öztopçu K (1999). *Elementary Azerbaijani*. Santa Monica & Istanbul: Ölmez.

Širaliyev M S & Sevortjan È V (eds.) (1971). *Grammatika azerbajdžanskogo jazyka*. Baku: Èlm.

Relevant Website

<http://www.turkiclanguages.com> – Website with many Turkish language resources.

This page intentionally left blank

B

Bactrian

P O Skjærø, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Bactrian was the local Iranian language of the Greco-Bactrian (or Kushana) kingdom in northern Afghanistan, founded by soldiers of Alexander the Great. The language is known from coins, a few stone and wall inscriptions (private and royal, the earliest from the 2nd century), and a small number of manuscript fragments from Turfan, as well as a large number of economic and legal documents, mainly on parchment, from northwestern Afghanistan dated between 342 and 781 A.D. It shares features with both Parthian, its western neighbor, and Chorasmian and Sogdian, its northern neighbors.

Bactrian is the only Iranian language written in Greek script. One letter was added to write *š* (similar in form to the Old Norse letter *þ*, which is commonly used to transcribe it, e.g., *κxνηþko* = *kanēško*). The letter <o> spelled *u* between consonants, *h* after vowels, and was probably not pronounced in final, but served as an end-of-word marker. This final *-o* often becomes *-a* before enclitics (e.g., *abo* ‘to,’ but *aba-fago* ‘to you’; *oto* ‘and,’ but *ota-kaldo* ‘and when’). Other final vowels are rare (except in the oldest inscriptions), and no words end in consonants. The consonants <s> and <z> may be ambivalent, as they correspond to both <s> and <z> and palatal <ś> and <ž> in the text in Manichean script.

The inscriptions are written in capital letters (without spaces between the words), while secular documents are written in a cursive ductus, in which several letters are sometimes identical. There are several Manichean or Buddhist texts in Greek cursive and one manuscript leaf in the Manichean script.

Gender (MASC-FEM) is distinguished in the definite article and in some adjectives (e.g., **torosaggo* [tursāng] ‘Turkish,’ FEM *torosanzo* [tursānz]) and in the perfect participles (e.g., *nabixt-igo* [nabixt-ig] MASC ‘written,’ FEM *nabixt-iso* [nabixt-is]). In the earliest inscriptions, there is still a two-case (direct and oblique) system of the noun, which in the documents

survives mainly in pronouns. Thus, in the inscriptions we find SING. *i bago* ‘the god.SING. DIR’ and PL. *i bage* ‘the god.PL.DIR’ as subject, but *bag-ano* ‘god-PL.OBL’ as genitive; *kanēško* ‘Kanishka. SING.DIR’ as subject, but *kanēški/kanēške* ‘Kanishka. SING.OBL’ as agent and genitive.

A definite animate direct object is indicated by the preposition *abo* ‘to.’

The verbal system is of the common Iranian type. There are three stems: present, past, and perfect (perfect participle = past stem + suffix *-igo*, FEM *-iso*; e.g., PRES *nabis-* ‘write,’ PAST *nabixt-*, PERF MASC *nabixt-igo*). Special features include modal forms formed from the indicative plus the original modal third singular ending.

optative: *ma froxoš-ond-ēio* [fraxwaš-und-ēy]
lest leave-INDIC.3RD.PL-OPT.3RD.SING
‘lest they abandon’

subjunctive: *boo-ado* [buw-ād]
become-SUBJ.3RD.SING
‘(that) he shall become’

and
boo-ind-ado [buw-ind-ād]
become-INDIC.3RD.PL-SUBJ.3RD.SING
‘(that) they shall become’

The perfect is formed with the old participle in **-aka-*

ot-ēia ... pidgirbo fromado kirdi eim-oano bag-ano
ki-di m-aska nibixt-ig-endi
and-he.OBL image ordered.PAST do.INF these-PL.OBL
god-PL.OBL REL-PART the.OBL-above written-PERF-PART-
COP.PRES.3RD.PL
‘and he ordered images to be made of these gods
which are written above’.

In the ergative construction, the relatively widespread phenomenon of letting verbs such as ‘give’ agree with the indirect object is found in Bactrian as well.

od-omo ladd-ēi iōgo zino
and-I.OBL give.PAST-be.2ND.SING one woman
‘and I have given you a (certain) woman’

A feature unusual in Iranian is the preposed negation in past tenses.

ko-ado-mēno n-isto paralado

that-PARTICLE-WE.OBL not-IS.3RD.SING sold.PAST
'that we have not sold'

A typically Bactrian construction is that of the subjunctive or optative with the particle *-an* used to express future eventuality.

- (1) *asid-ano oalo šatar-ano* [*šatar* < *šado* + *-tar*]
kald-ano abo to xoēo xoado lrogo oēn-ano
but-PART there happy.COMP-COP.SUBJ.1ST.SING when-
PART DO you lord self healthy see.PRES-
SUBJ.1ST.SING
'but I shall be happier there when I see you myself healthy'
- (2) *ot-ēio pido asagg-e iθo oilirdo at-ano abo ma lizo*
faro karano abo ma gao-ēio
and.PART-he.OBL on stone-PL thus arrange.PAST so
that-PART in DEF citadel for people water NEG
lack-OPT.3RD.SING
'and on (it) he placed stones so that in this citadel water might not be lacking for the people'

The particle *-do* is commonly attached to initial conjunctions, as in *kal-do* 'when,' *aki-do* 'who,' and

asi-do 'which'; the common form *oto* 'and' is from *odo* 'and' + *-do*.

Bibliography

- Humbach H (1966–1967). *Baktrische Sprachdenkmäler* (2 vols). Wiesbaden: Harrassowitz.
- Lee J & Sims-Williams N (2003). 'The antiquities and inscription of Tang-i Safedak.' *Silk Road Art and Archaeology* 9, 159–184.
- Sims-Williams N (1988). 'Bactrian language.' In Yarshater E (ed.) *Encyclopaedia Iranica*, vol. 3. London: Routledge and Kegan Paul. 344–349.
- Sims-Williams N (1989). 'Bactrian.' In Schmitt R (ed.) *Compendium linguarum Iranicarum*. Wiesbaden: Reichert. 230–235.
- Sims-Williams N (2000). *Bactrian documents from Northern Afghanistan I: Legal and economic documents*. Oxford and New York: Nour Foundation in association with Azimuth Editions and Oxford University Press.
- Sims-Williams N (2001 [2005]). 'Bactrian legal documents from 7th and 8th century Guzgan.' *Bulletin of the Asia Institute* 15.

Balinese

I W Arka, Australian National University, Denpasar
Bali, Indonesia

© 2006 Elsevier Ltd. All rights reserved.

Balinese (Bali) is an Austronesian language spoken by some 3 million people, mainly in the islands of Bali and Nusa Penida, Indonesia, but also in western Lombok and in transmigration sites in Lampung (Sumatra) and central Sulawesi. There is a general consensus that Balinese is a member of the Bali-Sasak-Sumbawa subgroup (Esser, 1938; Dyen, 1965; Mbete, 1990), but it is also seen as a member of a wider subgroup that includes Javanese (Blust, 1985).

History and Sociolinguistics

Balinese has had a literacy tradition for over a millennium. The earliest known Old Balinese (OB) texts are inscriptions on copper plaques dated to 882 C.E., concerning royal decrees (Goris, 1954). OB is characterized by the influence of Old Javanese (Kawi) and Sanskrit, which suggests the existence of cultural and language contact between Javanese and Balinese prior to the 9th century. Javanese influence

on Balinese intensified in the 14th–15th century, when Bali was controlled by the Javanese Majapahit Kingdom. Old Javanese elements and Sanskrit borrowings began to spread from highly formal – i.e., royal and religious – usage to everyday speech. These helped form the diglossic speech-level system of Modern Balinese, which is absent in OB (see Clynes, 1989, 1995). The speech-level system is invoked by differences in status between speech participants. As shown in **Table 1**, the English 'I' corresponds to several Balinese first person pronouns, each with a different specification of the speakers' and/or addressees' social

Table 1

| Pronominal forms | Relevant social information of the participants | |
|------------------|---|---------------|
| | Speaker | Addressee |
| <i>nira</i> | god | - |
| <i>gelah</i> | royal | - |
| <i>titiang</i> | - | highest caste |
| <i>tiang</i> | - | medium caste |
| <i>icang</i> | low caste | low caste |
| <i>kai</i> | - | nonhuman |

status, originally based on the traditional caste stratification (Arka, 1998); e.g., *icang* is used when both the speaker and the addressee are low-caste persons. While all of them are still in use now, *tiang* is widely used for polite first person irrespective of the caste of the addressee.

There are quite a large number of words like those in Table 1 in other categories, such as nouns, verbs, adjectives, prepositions and adverbs. They must be individually learned, because the related words are expressed by suppletive forms. The richness of the speech-level system is significant for Balinese verbal arts and linguistic politeness. However, the speech-level system is absent in the Bali Aga or Mountain Balinese (MB) dialect, suggesting that MB is a conservative dialect. Further evidence for this comes from the fact that MB (e.g., in the dialect of Sembiran) retains the Austronesian pronominal *aku* and *engko* and their corresponding bound forms *-ku* and *-mu*. These forms have disappeared in modern Lowland Balinese (LB). LB consists of several dialects showing phonological and lexical variations (Bawa, 1983), with Buleleng and Klungkung varieties being considered representative of standard modern Balinese.

Orthography and Phonology

The traditional Balinese script developed from the Old Javanese script, which itself originated from southern India. It is a syllabic system: a character represents a default CV (Consonant Vowel) syllable with V being phonetically [a] as in Table 2. Any specific opposition is indicated by a diacritical character on top of, below, before, and/or after it, as shown in Table 3. The line, as in the Roman script, runs from left to right.

While modern orthography in Roman script is also now commonly used, especially in paper writing, the

traditional script is the only script used in *lontar* (palm leaf) writing. *Lontar* writing and the tradition of *lontar* chanting, called *ma(be)basan*, are still practiced nowadays, primarily for religious purposes. Indonesian, not Balinese, is used as a medium of instruction in schools in Bali. However, Balinese, with its traditional script in paper writing and reading, is taught in primary and secondary schools.

Modern Balinese has six vowels, as shown in Table 4. Conventionally, the orthography *e* represents mid front [e] and the central [ə], e.g., *penek* [pənek] ‘climb’. Word-final grapheme *a* is pronounced [ə], e.g., *bapa* [bapə] ‘father’, but as [a] elsewhere, e.g., *bapanne* [bapanne] ‘his father’. VV sequences are not diphthongs but are treated as two syllables (Clynes, 1995), possibly with an intervocalic glide in certain dialects, e.g., *liu* [liu]~[liju] ‘a lot’.

Eighteen Balinese consonants are shown in Table 5. Word-final /k/ may be also alternatively realized by a glottal [ʔ] in certain dialects, but a glottal stop is not phonemic in Balinese.

Balinese allows a maximally C₁C₂VC₃ syllable structure, where only V is obligatory and C₂ is restricted to a liquid/glide, e.g., *alih* (V.CVC) ‘search’, *kranjang* (CCVC.CVC) ‘basket’, and *meme* (CV.CV). Stress is on the final syllable of a root, and a bound morpheme does not generally attract stress, particularly in the Badung dialect, e.g., *jemak* [dʒə.mak] ‘take’, *jemaka* [dʒə.mak.ə] ‘be taken’, and *jemakang* [dʒə.mak.əŋ] ‘be taken for’ (stressed syllables are underlined).

Morphosyntax

Balinese is an agglutinating language with relatively rich verbal and nominal morphology. A typical verbal expression involves a root and a voice morphology, which can be: (i) the homorganic nasal prefix N-, indicating an ‘active’ or ‘agentive’ voice, (ii) a zero prefix, indicating undergoer or objective voice, (iii) the middle (intransitive) voice prefix *ma-*, which expresses a wide range of meanings, e.g., reciprocal (*madiman* ‘kiss each other’), reflexive (*mapayas* ‘dress oneself’), agentive (*magae* ‘work’), patientive (*makeplug* ‘explode’), and stative-passive (*maadep* ‘be sold’). A verb may also have a causative or applicative affix. The applicative suffix *-in* is typically

Table 2

| | | | | | |
|----|------|----|----|----|-----|
| ᮊ | (h)a | ᮊ | ta | ᮊ | ba |
| ᮃ | na | ᮃ | sa | ᮃ | nga |
| ᮆ | ca | ᮆ | wa | ᮆ | pa |
| ᮉ | ra | ᮉ | la | ᮉ | ja |
| ᮊᮁ | ka | ᮊᮁ | ma | ᮊᮁ | ya |
| ᮊᮃ | da | ᮊᮃ | ga | ᮊᮃ | nya |

Table 3

| | | | | | | | |
|----|------|-----|------|-----|-------|-----|-------|
| ᮊ | [na] | ᮊᮁ | [ni] | ᮊᮃ | [nu] | ᮊᮅ | [ne] |
| ᮊᮇ | [nə] | ᮊᮇᮁ | [no] | ᮊᮇᮃ | [nar] | ᮊᮇᮅ | [nur] |

Table 4

| | Front | Central | Back |
|------|-------|---------|------|
| High | i | | u |
| Mid | e | ə | o |
| Low | | a | |

Table 5

| | | Labial | Alveolar | Palatal | Velar | Glottal |
|-----------|--------|--------|----------|---------|-------|---------|
| Stop | V-less | p | t | | k | |
| | V-ed | b | d | | g | |
| Aff. | V-less | | | tʃ | | |
| | V-ed | | | dʒ | | |
| Nasal | | m | n | ɲ | ŋ | |
| Fricative | | | s | | | h |
| Trill | | | r | | | |
| Lateral | | | l | | | |
| Glides | | w | | j | | |

associated with a locative or source role, whereas the applicative *-ang* is generally associated with a theme, goal/benefactive, or instrumental role; hence the contrast of *jemak* ‘take’ → *jemak-in* Y ‘take something from Y’ vs. *jemak-ang* Y ‘take something for Y’. The causative *-in* or *-ang* commonly appears with intransitive bases, but certain transitive verbs may have it, e.g., *diman* ‘kiss’ → *diman-ang* ‘make X kiss Y’. When *-in* and *-ang* appear with the same intransitive base, the derived verb generally contrasts in meaning, e.g., *tegak* ‘sit’ → *tegak-in* ‘sit on something’ (applicative *-in*) vs. *tegak-ang* ‘make somebody sit’ (causative *-ang*), *paek* ‘near’ → *paek-in* ‘move close(r) to something’ (applicative *-in*) vs. *paek-ang* ‘make something close(r) to something’ (causative *-ang*).

Word order is typically S(ubject)–V–O(bject), with S possibly coming after VO. In a double-object construction, the order of the two objects is fixed: S–VO_{Goal}–O_{Theme}.

Balinese appears to have symmetrical objects (Arka, 1998, 2003): either O_{Goal} or O_{Theme} could generally alternate to become S in a nonagentive voice construction, given the right context and intonation contour.

Balinese grammar has been well researched, mainly in the form of Ph.D. dissertations. Hunter (1988) and Beratha (1992) were historical-descriptive in perspective; Artawa (1994) was typological, highlighting the ergativity in Balinese syntax; and Clynes (1995) was also descriptive, focusing on Balinese phonology and morphosyntax (based on the dialect of Singaraja). Pastika (1999) was functional, focusing on the voice selection in Balinese narrative discourse. Arka (1998, 2003) was typological and theoretical, focusing on topics such as phrase structures, argument structures, and (reflexive) binding from a Lexical Functional Grammar (LFG) perspective. Wechsler and Arka (1998) and Wechsler (1999) were theoretical, from a Head-driven Phrase Structure Grammar (HPSG) perspective. Previous work on Balinese, not in the form of dissertations, also consisted essentially of descriptive sketches of grammar, e.g., Kersten (1970), Barber (1977), and Oka Granoka *et al.* (1985).

Dictionaries include Balinese-Indonesian (Warna *et al.*, 1993; Kersten, 1984; Ananda Kusuma, 1986), Indonesian-Balinese (Bahasa, 1975; Ananda Kusuma, 1986; Sutjaja, 2004), Balinese-English (Shadeg, 1977; Barber, 1979; Sutjaja, 2000), English-Balinese (Sutjaja, 2000), Kawi-Balinese-Dutch (Van Der Tuuk, 1897), and monolingual Balinese (Simpfen, 1985; Sutjaja, 2003).

Bibliography

- Ananda Kusuma S R (1986). *Kamus bahasa Bali*. Denpasar: CV Kayumas.
- Arka I W (1998). From morphosyntax to pragmatics in Balinese. Ph.D. diss., University of Sydney.
- Arka I W (2003). *Balinese morphosyntax: a Lexical-Functional approach*. Canberra: Pacific Linguistics.
- Artawa K (1994). Ergativity and Balinese syntax. Ph.D. diss., La Trobe University.
- Bahasa & Balai Penelitan (1975). *Kamus Indonesia-Bali*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Barber C C (1977). *A grammar of the Balinese language* (2 vols). Aberdeen: Aberdeen University.
- Barber C C (1977). *A Balinese-English dictionary*. Aberdeen: Aberdeen University.
- Bawa I W (1983). Bahasa Bali di daerah propinsi Bali: sebuah analisis geografi dialek. Ph.D. diss., Universitas Indonesia.
- Beratha N L S (1992). Evolution of verbal morphology in Balinese. Ph.D. diss., Australian National University.
- Blust R (1985). ‘The Austronesian homeland: a linguistic perspective.’ *Asian Perspectives* 26, 45–67.
- Clynes A (1989). Speech styles in Javanese and Balinese: a comparative study. M.A. thesis, Australian National University.
- Clynes A (1995). Topics in the phonology and morphosyntax of Balinese. Ph.D. diss., Australian National University.
- Dyen I (1965). ‘A lexicostatistical classification of the Austronesian languages.’ *International Journal of American Linguistics*, Memoir 19.
- Esser S J (1938). ‘Languages.’ In *Atlas van Tropisch Nederland*. Batavia: Koninklijk Nederlandsch Aardrijkskundig Genootschap. Sheet 9b.

- Goris R (1954). *Prasasti Bali*. Bandung: Lembaga Bahasa dan Budaya.
- Hunter T M (1988). Balinese language: historical background and contemporary state. Ph.D. diss., Michigan University.
- Kersten J (1970). *Tatabahasa Bali*. Flores: Arnoldus.
- Kersten J (1984). *Bahasa Bali*. Ende-Flores: Nusa Indah.
- Mbete A M (1990). Rekonstruksi protobahasa Bali-Sasak-Sumbawa. Ph.D. diss., Universitas Indonesia.
- Oka Granoka I W, Udara Naryana I B & Jendra I W (1985). *Tata bahasa Bali*. Denpasar: Proyek Pengembangan Bahasa dan Sastra Indonesia dan Daerah Departemen Pendidikan dan Kebudayaan.
- Pastika I W (1999). Voice selection in Balinese narrative discourse. Ph.D. diss., Australian National University.
- Shadeg N (1977). *A basic Balinese dictionary*. Denpasar: Dharma Bhakti.
- Simpén I W (1985). *Kamus Bahasa Bali*. Denpasar: PT Mabhakti.
- Sutjaja I G M (2000). *Practical Balinese-English-English-Balinese dictionary*. Denpasar: Bali Post.
- Sutjaja I G M (2003). *Kamus sinonim bahasa Bali*. Denpasar: Lotus Widya Suari/Universitas Udayana.
- Sutjaja I G M (2004). *Kamus Indonesia-Bali-Inggris*. Denpasar: Lotus Widya Suari/Universitas Udayana.
- Van Der Tuuk H N (1897). *Kawi-Balineesch-Nederlansch Woordenboek*. Batavia: Landsdrukkerij.
- Warna I W *et al.* (1993). *Kamus Bali-Indonesia*. Denpasar: Dinas Pendidikan Dasar Propinsi Dati I Bali.
- Wechsler S (1999). 'HPSG, GB, and the Balinese Bind.' In Kathol A, Koenig J-P & Webelhuth G (eds.) *Lexical and constructional aspects of linguistic explanation*. Stanford: CSLI. 179–195.
- Wechsler S & Arka I W (1998). 'Syntactic ergativity in Balinese: an argument structure based theory.' *Natural Language and Linguistic Theory* 16, 387–441.

Balkans as a Linguistic Area

V A Friedman, University of Chicago, Chicago, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Definitions

Sprachbund

Among the proposed glosses for *sprachbund* are 'linguistic league', 'linguistic area', 'convergence area', and 'diffusion area', but here I will treat *sprachbund* as a loanword into English, like the French *genre*, so henceforth it will be neither capitalized nor italicized. In modern terms, a sprachbund is understood as two or more geographically contiguous and genealogically different languages sharing grammatical and lexical developments that result from language contact rather than a common ancestral source. (Some linguists set the minimum number at three, but I would argue that the convergent and diffusion processes consitutive of a sprachbund are the same for two languages as for three.) In his original formulation of the concept, first in 1923 in a Russian journal article and again in 1928 at the first International Congress of Linguists, N. S. Trubetzkoy used Bulgarian as his example of a language that belongs to the Slavic linguistic family and at the same time to the Balkan sprachbund. In the case of the Balkan sprachbund, the languages are in fact all Indo-European (excluding Balkan Turkic), but they belong to groups that were separated for millennia, and thus, upon coming

back into contact, had become sufficiently distinct for contact phenomena to be distinguished from inherited phenomena.

Balkan

The use of the term 'Balkan' (from Turkish, *balkan* 'forested mountain', also the name of a mountain range in Central Bulgaria) to refer to the peninsula also known as Southeastern Europe dates from the 19th century, when European attention turned to Ottoman Turkey, which then included most of what became the Balkan states. As a geographic entity, the Balkan peninsula is unproblematically defined on three sides as the land mass defined by the Adriatic, Mediterranean, and Black Seas, but the northern geographic boundary cannot be set in any nonarbitrary way that is applicable without qualifications in terms of either politics or linguistics. In modern geopolitical terms, from the 1920s to 1991, the Balkans were most frequently understood as comprising Albania, Bulgaria, Greece, Romania, Turkey in Europe, and former Yugoslavia.

The Balkan Languages

For linguistics, the Balkan sprachbund has traditionally consisted of Albanian, Greek, Balkan Romance (BR), and Balkan Slavic (BS). Albanian is divided into two dialects, Gheg north of the river Shkumbi and Tosk south of it. The modern standard is based on

northern Tosk. Mainland Greek is also divided between northern and southern dialects at the Gulf of Corinth and the northern frontier of Attica, the southern dialects of the Peloponnese being the basis of the standard vernacular Dhimotiki. During the 19th century, Modern Greek was still called *Romaic*, i.e., ‘Roman’, a reference to Byzantium as the second Rome. BR consists of Romanian, Aromanian, Megleno-Romanian (MR), and Istro-Romanian. Dalmatian, a remnant of West Balkan Romance, whose last speaker died in 1898, is rather poorly attested and generally does not figure in Balkan linguistic accounts. Istro-Romanian is, like Arbëresh (the Albanian of Italy) and Asia Minor Greek (until the exchange of populations between Greece and Turkey in 1923), outside the Balkan geolinguistic area (see ‘Balkan Languages vs. Languages of the Balkans’). The Romanian standard is based on the Wallachian dialects of the south, as is the standard of the Republic of Moldova, which at various times has called its official language Moldovan or Romanian. (At present [31 October 2004] the official name is Moldovan.) Aromanian, spoken in Albania, Greece, the Republic of Macedonia, and southwestern Bulgaria (with a large diaspora in Romania, especially Dobrogea) is divided into north/west dialects of Albania and western Macedonia and south/east dialects of Greece and eastern Macedonia. A standard based primarily on the eastern dialect is in use in the Republic of Macedonia. MR survives in seven villages near Gevgelija in the southeast of the Republic of Macedonia and across the border in Greece. During the 19th century, BR was often called Wallachian. The term ‘Vlah’ can be used as a convenient cover term for BR south of the Danube (Aromanian plus Megleno-Romanian). BS consists of Bulgarian, Macedonian, and the southeast Serbian (Torlak) dialects. Bosnian/Croatian/Serbian (BCS) together with Slovene, form the West South Slavic group, and Macedonian and Bulgarian comprise East South Slavic. The Bulgarian standard is based on its eastern dialects, the Macedonian standard on its west-central dialects. The northern and western boundaries of Torlak as a Balkan dialect are variously defined using phonological or morphological criteria. The narrowest definition is morphological, e.g., the isogloss for the presence of the postposed definite article; the broadest definition is phonological, e.g., the absence of distinctive vocalic length and tone. During the 19th century, BS was often called ‘Bulgarian,’ and Bulgarian and Serbian linguists and armies fought over where to draw a line between Bulgarian and Serbian. Unable to adjust to modern times, many Bulgarian linguists still cling to the 19th-century practice.

Romani Despite having been summarily dismissed by traditional Balkan linguists, Romani in the Balkans displays many of the same contact-induced structural phenomena and is increasingly present in Balkanological works. Two of the four main dialectal groups of Romani are spoken in the Balkans: Balkan and Vlax (not to be confused with Vlah). The Vlax dialects of Romani take their name from the fact that they took shape in Romania, but they are now dispersed all over Europe and beyond. In the Republic of Macedonia, a Romani standard is emerging on the basis of the Arli dialect of the Balkan group. Unless otherwise specified, references to Romani refer to those dialects spoken in the Balkans.

Turkish Balkan Turkish is divided into two major dialect groups: West Rumelian Turkish (WRT) and East Rumelian. The boundary between the two corresponds roughly to the east-west line of Bulgarian dialects. The Christian Gagauz of Bulgarian and Romanian Dobrudja and Gagauz Yeri in Moldova and adjacent parts of Ukraine speak a language in the Oghuz group – to which Turkish also belongs – which was recognized as official in the USSR in 1957. Although most Balkan linguistic studies treat Turkish as an adstratum, contributing lexicon and phraseology but very little else (aside from evidentiality, see ‘Evidential’ below), WRT and Gagauz also participate to a certain extent in the Balkan sprachbund. Most of Gagauz, however, ended up in the former Russian Empire, due to migration and border changes. As a result, most of Gagauz is now more influenced by Russian, while the dialectal Gagauz remaining in the Balkans is in need of description.

Jewish Languages Judezmo, the language of the Jews expelled from Spain in 1492, became the majority language among Balkan Jews, overwhelming Judeo-Greek (Yavanic, Yevanic), which survived in the Romaniote liturgy and some enclaves in Epirus. (A written version of Judezmo based on literal translation from Hebrew is known among scholars as Ladino.) Although most speakers of both Judezmo and Judeo-Greek were murdered in the Holocaust, these languages survive as endangered languages and also participated in Balkan linguistic processes.

Balkan Languages vs. Languages of the Balkans There are many other languages spoken in the Balkans in enclaves with varying social relations, e.g., Armenian, Circassian (until 1999), German, Hungarian, Ruthenian, Tatar, Ukrainian, Yiddish, etc. Aside from the dialects spoken in Romania, most of these are outside the geolinguistic Balkans, which for

our purposes has a northwest boundary defined by contiguous Albanian dialects that join the major Torlak isoglosses continuing to the Danube. (Such a definition includes the southernmost Slavic dialects of Montenegro as well as the Slavic dialects of northern Kosovo, neither of which fall in the Torlak group. In terms of the Balkan sprachbund, these dialects do show some important transitional features, which will be noted.) For the most part, the enclave languages were late arrivals or outside the area of intensive diffusion/convergence and did not participate in the type of complex Balkan multilingualism that characterizes the sprachbund as a whole. We can thus distinguish Balkan languages, i.e., those in the sprachbund, from languages of the Balkans, i.e., languages spoken in the Balkan peninsula.

History of Balkan Linguistics

1770–1861

The earliest collections of Balkan linguistic material were intended to eliminate Balkan linguistic diversity. The 1770 Greek-Aromanian-Albanian vocabulary of T. Kavaliotis and the 1793 or 1794(?) Greek-Aromanian-Macedonian-Albanian lexicon of Daniil of Moschopolis (Albanian Voskopoja) were explicitly aimed at the Hellenization of the speakers of other Balkan languages. The first was republished in 1774 by J. Thunmann, who was the first to suggest that Albanians and Romanians were descended from Illyrians, Dacians, and Thracians, thus laying the groundwork for the substratum theory of Balkan linguistics. The second was republished in 1814 by M. Leake, who suggested that similarities among Albanian, BR, and Greek were due to BS influence. His one concrete example was the postposed definite article. It was this same phenomenon that most impressed J. Kopitar, whose 1829 characterization of BR, BS, and Albanian as *drey lexikalisch verschiedenen, aber grammatisch identischen Sprachen* ‘three lexically distinct but grammatically identical languages’ – which he attributed to the influence of a Thracio-Illyrian substratum – is taken as the earliest formulation characterizing the Balkan sprachbund. Kopitar also noted the replacement of infinitival with subjunctive constructions and the formation of the future using ‘want’ as shared with Greek and Serbian as well.

A. Schleicher is sometimes cited as the first to formulate the Balkan sprachbund in 1850, when he writes of Albanian, BR, and BS saying *eine Gruppe aneinandergränzender Sprachen zusammengefunden hat, die bei stammhafter Verschiedenheit nur darin*

übereinstimmen, dass sie die verdorbensten ihrer Familie sind (‘a group of propinquitous languages has coalesced that, being of different lines of descent, agree only in the fact that they are the most corrupt in their families’). However, since he gives no indication of the causes of this ‘corruption’, his formulation differs from Kopitar’s mainly in its ideology of language change as degeneration.

The next real advance in the development of Balkan linguistics was F. Miklosich’s 1861 article on Slavic elements in Romanian, which added genitive-dative merger (see ‘Genitive-Dative Merger’), object pronoun doubling (see ‘Resumptive Clitic Pronouns [Reduplication, Replication]’), and the formation of teens (see ‘Numeral Formation: The Teens’). Miklosich accorded more attention to Greek and was also the first to adduce a number of phonological changes, including the development of stressed schwa (see ‘Vowel Reduction and Raising’) and the raising of unstressed /a/ and /o/ to schwa and /u/, respectively (see ‘Stressed Schwa’).

1861 Onward

The next six decades were characterized by the gathering of materials relating to specific Balkan languages or specific aspects of individual or pairs of Balkan languages. The 1920s saw the basic syntheses and theoretical formulations that continue to inform the field. Trubetzkoy’s contribution has already been described. In 1925, A. Seliščev attempted a balanced account of Turkish, Slavic, Latin, Greek, and substratum languages as the sources of various Balkanisms, i.e., the similarities among the Balkan languages that can be attributed, at least in part, to shared, contact-induced change. Sandfeld (1930) tried to attribute almost all the commonalities of the Balkan sprachbund to the influence and prestige of Byzantine Greek. Other scholars have laid particular emphasis on Balkan Latin as the primary causal factor, while our knowledge of the pre-Latin non-Hellenic languages of the Balkans remains too meager for almost any serious speculations beyond the lexicon.

While the 1920s saw the establishment of Balkan linguistics as a subdiscipline within linguistics, the period from 1930 to 1960 was characterized by slow growth and was also the period when the insights gained in Europe finally came to the attention of North American linguists. From the 1960s onward, there has been a constant increase in the production of studies pertaining to the Balkan languages and Balkan linguistics. At the same time, studies of such contact-induced phenomena as creolization, code switching, and language shift have led to the

identification of contact linguistics as an overarching field of study. More recently, in the past decade or so, a renewed interest in linguistic typology has brought forward questions of the extent to which the Balkan sprachbund is or is not part of a larger European linguistic area, defined more by typological profile without necessarily identifying specific paths of diffusion or convergence. We will return to the question of Eurology vs. Balkanology in 'Causation'.

Balkanisms

This section surveys some of the principal Balkanisms (see '1770–1861') as identified during the course of the past two centuries. Although system, not mere inventory, must be the basis of detailed study, and a given surface phenomenon may function differently in different systems, it is nonetheless convenient to use lists as a kind of shorthand for the systemic relations that can yield the most insights. We do not want to fetishize the labels for these systemic manifestations, assigning numeric values to them and tallying up the number of points a language 'scores'. Rather these labels stand for complex interrelations that include differences as well as similarities that must be elucidated in their larger contexts (cf. Friedman in Reiter, 1983).

Phonology

In contradistinction to linguistic areas such as the Caucasus, the Northwest Coast, and South Asia, where phonological features such as glottalization and retroflexion are among the most salient commonalities, there are no truly pan-Balkan phonological features. Rather, there are articulatory tendencies of greater or lesser extent.

Vowel Reduction and Raising The reduction of unstressed vowels to schwa or nonsyllabic elements (and thence sometimes to zero) as well as the raising of unstressed mid-vowels (/e/ and /o/) to high vowels (/i/ and /u/, respectively) can be treated as Balkan, albeit not pan-Balkan. Both Albanian and BR show a tendency to reduce unstressed vowels as early as the Latin period, e.g., Lat. *imperator* > Albanian *mbret* and Romanian *împărat* 'king'. While shared phonological tendencies in Albanian and BR, like shared vocabulary of pre-Latin origin, are attributed by some scholars to substrate influence, the evidence of vowel reduction in Western Romance leads other scholars to suggest that this is a typological rather than an areal feature. Nonetheless, the raising and/or elimination of unstressed vowels is characteristic of southeastern Macedonian, eastern Bulgarian,

northern Greek, BR, and Gheg, although the details differ among these languages.

Stressed Schwa All the Balkan languages and their dialects possess the classic European five vowel system /a, e, i, o, u/, at least under stress. A phenomenon common in the Balkans is the existence of a stressed schwa, but its status as a contact-induced phenomenon is not pan-Balkan. Greek lacks stressed schwa altogether. In Macedonian, almost all the dialects outside the west-central area have stressed schwa, but of different origins in different areas, and some western peripheral dialects also lack stressed schwa. Most of Bulgarian has stressed schwa, but not the Teteven-Erkech and central Rhodopian dialects. In Albanian, stressed schwa develops from nasal *â* only in Tosk, but it is incorrect to characterize all of Gheg as lacking stressed schwa, since it also occurs in central Gheg as a result of later processes. Romani has schwa when in contact with languages that have it. WRT has a tendency to lower and front the high back unrounded vowel to schwa.

Other Vowels Most Balkan languages lack front rounded vowels, but most of Albanian has /y/, or, in West Central Gheg, /ö/. Southern Montenegrin dialects in contact with Albanian also have /y/, but East Central Gheg, which is mostly in Macedonia, unrounds /y/ to /i/, as does southernmost Tosk (Lab, Çam, Arvanitika), in contact with Aromanian and Greek (which also merged /y/ with /i/, a change that had not yet been completed in the 10th century). Similarly, WRT tends to eliminate /ö/ by merging it with /o/ or /ü/ (more rarely /e/), and /ü/ (like /u/ and /i/) becomes /i/ word finally. Other vocalic phenomena that have been suggested are relatively localized.

Consonants The alternation of clear /l/ before front vowels and velar /l/ elsewhere is characteristic of BS (including Torlak but not the rest of BCS), Northern Greek, Balkan Romani, and Vlah, but not Albanian, where the two sounds are in phonemic contrast, nor Daco-Romanian and Southern Greek, where only clear /l/ occurs. Aromanian has Greek and Albanian interdental and Greek voiced velar and palatal fricatives in loanwords from Albanian and Greek, but these tend to be replaced by corresponding stops and the palatal glide by speakers who do not know Greek or Albanian, particularly the younger generation in Macedonia.

Aside from Greek, most Balkan languages have an opposition between strident palatal affricates, on the one hand, and mellow palatals, dorso-palatals, or palatalized velars, on the other. The opposition is neutralized in Albanian, BS, and WRT dialects in

Kosovo, parts of Western Macedonia, and along the Serbo-Bulgarian border. Northern Greek has palatals lacking in the south.

In western Macedonia, the velar fricative is generally lost or replaced in Albanian, Macedonian, and WRT, a phenomenon that extends into parts of Kosovo, as well as adjacent Serbia, much of Montenegro, and Bosnia-Herzegovina, where the preservation of BCS /x/ is characteristic of Muslim and some Catholic dialects now Bosnian and Croatian, respectively.

In the northern Ghëg of Malësia e Madhe, final devoicing is a phenomenon shared with adjacent Montenegrin dialects. It is worth noting that final devoicing is atypical for most of the rest of BCS and Ghëg, and it appears rather to be a Macedonian feature extending into this region. Such influence also seems to be the case in the transitional Ghëg and northern Tosk dialects. Some of the Romani dialects in this region also have final devoicing, and in the WRT of these regions final devoicing, which is usually limited to stops in Turkish, extends to fricatives. Five of the seven MR villages also have final devoicing.

Prosody Although prosodic distinctions of length, and in some cases pitch, were present in the attested ancestors of the Balkan languages, the modern Balkan languages are generally characterized by the absence of length and tone and the presence of a stress accent that usually does not move further back in the word than the antepenultimate syllable. If stress does move further back, there is usually a secondary stress on one of the last three syllables. However, Northern Ghëg and Southern Tosk preserve Common Albanian length, and Southeastern Macedonian has new long vowels as the result of loss of intervocalic consonants and elision. Similar new long vowels occur in Gora, a string of Slavic-speaking Muslim villages along the western and northern slopes of Mounts Korab and Šar in northeastern Albania and the southwestern corner of Kosovo. The most significant isoglosses (fixed antepenultimate stress, postposed article, etc.) link Goran with the northwest Macedonian dialects rather than with the Serbian of Prizren.

Morphosyntax

Grammaticalized Definiteness In BS, BR, and Albanian, native demonstrative pronouns have been encliticized or suffixed to nominals (normally the first in the noun phrase) and become definite articles. The article follows a plural marker, if any, and in BS the clitic-like nature of the article is seen in that it does not trigger certain morphophonemic alternations,

e.g., Macedonian *starec* ‘old man’, *starci* ‘old men’ but *starecot* ‘the old man’ and not **starcot*. Hamp (1982) adduces evidence suggesting that the autochthonous language that became Latinized into Romanian and with which the ancestor of Albanian was in contact might already have had a postposed definite article by the time of contact with Latin. Common Slavic already had a postposed relative pronoun **ji* affixed to adjectivals to denote definiteness, as this phenomenon is attested in Old Church Slavonic (OCS; 9th–11th centuries), and the morphology (but not the grammatical meaning) survives in Slavic outside the geopolitical Balkans. Remnants of this older definite/indefinite opposition survive in West South Slavic adjectives, and traces of the morphology occur in BS, e.g., Macedonian *star* ‘old INDEF. MASC’, *stariot* ‘old DEF.MASC’, where the /i/ indicates that the newer definite article has been suffixed to a definite adjectival form. Scandinavian and dialectal North Russian also have postposed definite articles of pronominal origin, and Czech, which has been in close contact with German, has uses of its deictics that are basically articular. These typological parallels and historical antecedents, however, do not change the fact that the BS postposed definite article developed during the period of its contact with BR and Albanian.

Greek and Romani have preposed definite articles, both based on native material. In the case of Greek, the pronoun that became an article was still mostly demonstrative and was facultative except with proper names in Homeric, but it was obligatory in Attic. Romani articles look like borrowings from Greek, e.g., MASC NOM SG *o* FEM NOM SG *i*, but the oblique forms /le/ and /la/ in Vlax dialects demonstrate that the Romani articles are derived from native demonstratives, reflecting the regular change of **t > l*, which occurred prior to contact with Greek. It was contact with Greek, however, that triggered the transformation of native material into definite articles, and Romani usage patterns very much like Greek. Romani dialects outside the Balkans in contact with languages lacking definite articles tend to lose them.

The use of an atonic form of the numeral ‘one’ as an indefinite article is characteristic of the Balkan languages and, even though such developments are common in many languages, is arguably a Balkanism. ‘One’ was not used in this function in OCS, Ancient Greek, or Latin, but it was so used in Orkhon Turkic (8th century C.E.). To this we can add the fact that such usage does not occur in East Slavic. Usage in Turkish, Albanian, and BR is at a similar level of frequency to that of English, although details in individual grammars will cause some lack of isomorphism. Usage in

BS and in Greek is approximately half that of the other Balkan languages, while usage in Romani in the Balkans patterns with BS and Greek, and Romani elsewhere patterns like its contact languages. An indication that this is an areal phenomenon despite the occurrence of such usages in Western Europe and elsewhere is the fact that, as one moves north and east through West South Slavic territory, the usage becomes increasingly restricted.

Finally, we can mention here the phenomenon of double determination, i.e., the presence of a definite article on a noun modified by a demonstrative pronoun. Such usage occurs in Greek, BR, BS, Albanian, and Romani, although the rules and relative frequency and acceptability of the construction vary. In Greek it is obligatory, e.g., *autós o ánthrōpos* or *o ánthrōpos autós* but not **autós ánthrōpos* 'this person'. In Romanian, the article is not used if the deictic is preposed, but is used if it is postposed (and the deictic takes the so-called deictic particle *-a*): *omul acesta* but *acest om* 'this person', cf. Aromanian *aistă carte*, *cartea aistă* 'this book'. Megleno-Romanian has frequent double determination *tsista lup-u* 'this wolf-DEF', but indefinite nouns also occur *tsista drăc* 'this devil-INDEF'. In Albanian, the deictic is preposed to either the indefinite or definite: *ai njeri*, *ai njeriu* 'this person'. In BS, double determination occurs but is considered dialectal, Macedonian *ovoj čovekov* (vs. *ovoj čovek*) 'this person', or Torlak *taja starata* 'that old [lady]'. Romani permits but does not require the use of a definite article with a demonstrative, in which case the article must precede the substantive but the demonstrative can precede or follow: *kova manuš*, *kova o manuš*, *o manuš kova* 'this person'. Double determination or the order noun-determiner is pragmatically more thematic in the discourse.

Resumptive Clitic Pronouns (Reduplication, Replication) Balkan languages are characterized by the use of clitic or weak resumptive object pronouns that agree in gender, number, and case with the non-clitic/strong pronoun or substantive they refer to. This phenomenon is called (object/pronoun) reduplication/doubling in Balkan linguistics and is connected to expressions of definiteness, referentiality, and animacy: the first candidates for reduplication are personal pronouns (inherently definite and, in the first two persons, usually human), then indirect objects (usually human, often topicalized), then definite direct objects, and finally specific or topicalized direct objects.

From a morphosyntactic point of view, there are four types of reduplication: pronominal object doubling, substantival object replication, pronominal possessive doubling, and substantival possessive

replication. All four phenomena can be illustrated in the following Macedonian sentence:

| | | | |
|----------------------|--------------------|-----------------|---------------------|
| Tatko mi | moj | i | majka |
| <i>father me.DAT</i> | <i>my.M</i> | <i>and</i> | <i>mother</i> |
| mu | na car-ot | im | rekoa |
| <i>him.DAT</i> | <i>to king-the</i> | <i>them.DAT</i> | <i>said.3PL.AOR</i> |
| nim | da | mu | gi |
| <i>them.DAT</i> | <i>SP</i> | <i>him.DAT</i> | <i>them.ACC</i> |
| dadat | knigi-te | na | dete-to |
| <i>give.3PL.PRES</i> | <i>books-the</i> | <i>to</i> | <i>child-the</i> |

'My father and the king's mother told them to give the books to the child.'

The first three of these expressions are facultative and could be replaced by *tatko mi*, *majkata na carot* (*majka* is definite), and *im*, respectively. The reduplication serves to emphasize or focus the referent of the reduplicated pronoun. The last set of reduplications, *mu ... na deteto* and *gi ... knigite*, are obligatory in standard Macedonian and, for the most part, in the western dialects on which it is based. The norm requires reduplication for definite direct objects and all indirect objects. In practice, however, even the most normative grammar shows that specificity or topicalization rather than definiteness is the trigger (Koneski, 1967: 232):

| | | |
|-------------------------|--------------------|----------------------|
| kako vistinski ja | doživuvame | edna situacija |
| <i>how truly it.ACC</i> | <i>experience.</i> | <i>one situation</i> |
| | | <i>1PL.PRES</i> |

'how we actually experience a [given] situation'

Pronominal object doubling occurs in all of BS (and southern Montenegro), BR, Albanian, Greek, and Romani. It is conditioned by discourse factors such as emphasis or focus and can be compared to the use of subject pronouns. Just as the fact that the subject is marked on the verb makes the subject pronoun redundant unless there is a need for emphasis or specification, so, too, the clitic pronominal object, which is the required form if the object is a pronoun, makes the full form redundant except under similar discourse-bound circumstances. The absence of such doubling from the rest of BCS is a diagnostic separating Balkan from non-Balkan Slavic.

The clitic replication of oblique nominals shows how grammatical change can enter a language via discourse phenomena and at the same time supports Topolińska's observation that analytic markers of referentiality are characteristic of convergent development. Object reduplication is another scalar Balkanism. It is rare in Torlak and used only for emphasis and thus separates East from West South Slavic. Similar conditions hold for Romani except in possessive constructions. Object reduplication is more pragmatically conditioned and less grammaticalized

in Bulgarian, Romanian, and Greek, where the phenomenon signals topicalization, focus, or emphasis, and is restricted by factors such as animacy (or humanness) and degree of referentiality (definiteness, specificity, determinacy, etc.). In Albanian, Vlah, and West Macedonian, reduplication has become grammaticalized. It is most frequent in Macedonian, where, unlike in the other Balkan languages, it can even occur (facultatively) with indefinite indeterminate pronouns such as *nikoj* 'nobody'.

While it lacks a definite article, Turkish does have a special accusative marker used for definite or specified direct objects. The following proverb illustrates how the Turkish definite accusative is rendered by Balkan object reduplication. Note that Greek and Bulgarian have reduplication with an indefinite object, indicating its specificity:

| | | | |
|--|----------------|------------|-------------------|
| Turkish: Yavaş | baş | kılıç | kes-mez (Turkish) |
| gentle | head- | sword | cuts-not |
| | DEFACC | | |
| Bulgarian: Pokorena | glava | sabja | ne ja seče |
| bent | head | sword | not it.ACC cuts |
| Greek: Kefáli | proskynēméno | spathi | dhèn |
| head | bent | sword | not |
| tò | kóvei | | |
| it.ACC | cuts | | |
| Romanian: Cap-ul | plecat | nu l | taie sabia |
| head- | bent | not it.ACC | cuts sword. |
| DEF | | | DEF |
| Albanian: Kokën | e | falur | yatagan-i |
| head | PART.F.DEF.ACC | bent | sword-DEF |
| nuk | e | pret | |
| not | it.ACC | cuts | |
| 'A/The sword does not cut off a/the bent head' | | | |
| (= Keep your head down.) | | | |

Possessive doubling is a more restricted phenomenon. The use of dative clitics to indicate possession in Macedonian is limited to kinship terms, Aromanian has special possessive clitics that can only be used with kinship terms, and Albanian also has special possessive constructions for kinship terms. In Bulgarian, possession is usually signaled by a dative clitic following the definite form of the noun, and possessive adjectives, which are the norm in Macedonian, are more emphatic in Bulgarian. In Greek, clitic dative pronouns after the definite form of the noun is the normal manner of indicating possession, and emphasis is rendered by adding the appropriate form of the adjective *dikós* '[one's] own' immediately before the pronoun. However, pronominal doubling is also used colloquially for emphasis:

| | | | |
|-----------|--------|--------|--------|
| to | vivlio | mou | mena |
| the | book | me.GEN | me.GEN |
| 'my book' | | | |

Romanian also has such clitic doubling colloquially:

| | | |
|-------------------------|-----|---------------|
| propria-mi | mea | semnătura |
| own.FEM-me.DAT | my | signature.DEF |
| 'my very own signature' | | |

Substantival possessive replication occurs in all the Balkan languages, but the details differ from language to language. The Turkish construction of genitive possessed plus pronominal suffix on the possessor is the normal pattern:

| | |
|--|------------|
| kral-in | anne-si |
| king-GEN | mother-his |
| 'the king's mother = the mother of the king' | |

Genitive-Dative Merger Albanian, BS, BR, and Greek have no formal (i.e., surface) distinction between the shape of the genitive and the shape of the dative, the dative having replaced the genitive except in Greek, where the genitive replaced the dative. The same forms thus do double duty for marking possession and indirect objects. Romani and WRT maintain the genitive/dative distinction, and the situation is more complicated in Albanian and MR. Albanian has merged genitive and dative but has a distinct ablative. The dative is used as the object of a verb, the genitive is preceded by a particle of concord, and the ablative is the object of certain prepositions or in apposition to another substantive. In the indefinite plural, however, Albanian has a special ablative form in *-sh*. Pronominal declension also has a distinct ablative form used with certain prepositions, NOM *nga unë/djali* 'from me/the boy', ACC *për mua/djalin* 'for me/the boy', DAT *më tha mua/i tha djalit* 'he told me/the boy (with initial clitic reduplication)', ABL *prej meje/djalit* 'from me/the boy'. MR preserved a remnant of the genitive-dative distinction, albeit only in the speech of the oldest generation: *cari* 'who' *pe cari* 'whom.ACC', *la cari* 'to whom.DAT' but *al cruji* 'of whom, whose'. Elsewhere, the dative and accusative are distinct, and the genitive is identical to the dative.

Analytic Case Relations All the Balkan languages have simplified their inherited patterns of inflection. Eastern Macedonian and colloquial Bulgarian have gone the farthest, completely eliminating all traces of case morphology other than accusative personal pronouns and accusative vs. dative clitics. The marking of nonclitic dative objects is by means of the preposition *na* and the accusative pronoun. All other case relations are likewise indicated syntactically throughout BS, usually by a preposition but sometimes just by apposition. Western Macedonian preserves a distinctive set of dative synthetic pronouns, and, in the

dialects that serve as the basis for the standard, a few remnants of animate singular masculine accusatives. As one moves further to the periphery of BS in the southwest and north, the complexity of case marking increases to include feminine accusatives, masculine datives, feminine datives, and eventually, in Gora and Torlakia, oblique plurals. In the Torlak dialects and the Macedonian dialects around Korça in Albania, case marking also occurs in the definite article. The other Balkan languages all retain at least three distinct cases (nominative, accusative, and genitive-dative).

Balkan Romani and WRT both preserve their full inflectional systems, but with tendencies toward simplification that show an intersection between the areal and typological. From a typological point of view, it is the peripheral cases that are expected to be lost first, and this is precisely what happens. Thus, WRT exhibits dative-locative confusion:

gitti-k Selanik-te
went-1PL.AOR Salonica-LOC
'We went to Salonica'

There is also a tendency to eliminate case marking in locational postpositions:

ürti üsti [vs üstü-n-de] kedi-ler
blanket top top-its-LOC cat-PL
'on top of his blanket [there were] cats'

Romani dialects in contact with BS tend to replace the locative with the dative and the dative, locative, and ablative with prepositional constructions derived from case affixes, themselves of postpositional origin:

| | | | | | | | |
|-----------------|------------|----------------|--------------|------------|---|-----------------|------------|
| jekh-e | aindž-a-te | vs. | jekh-e | aindž-a-ke | > | k-i | jekh aindž |
| one- | field- | | one- | field- | | to-FEM | one |
| OBL | OBL- | | OBL | OBL- | | field | |
| | LOC | | | DAT | | | |
| 'in a field', | | | 'to a field' | | | 'in/to a field' | |
| aindž-a-tar | = | tar-i aindž | | | | | |
| field-OBL-ABL | | from-FEM field | | | | | |
| 'from a field'. | | | | | | | |

Outside the pronouns, a distinct Romani accusative is limited to animate (or in some dialects referential) nouns, while in Turkish accusative marking is limited to definite or specific direct objects (see 'Resumptive Clitic Pronouns' [Reduplication, Replication]).

The vocative survives in all the Indo-European Balkan languages, and some argue that this preservation is a shared archaism, reinforced by contact, which is consistent with the direct encounters that lead to contact phenomena. It runs counter to the tendency toward analytism, however.

Analytic Gradation of Adjectives Although the comparative is analytic in all the Balkan languages,

remnants of synthetic comparatives survive at the peripheries, i.e., Greek has a number of inflected comparative forms, and northern Torlak preserves a very limited set. In the rest of BS, analytic comparatives with *po* are realized with almost complete consistency. Southern Montenegrin dialects also have analytic adjectival gradation using the same markers. BR, Albanian, and most Balkan dialects of Romani have complete consistency in the analytic marking of the comparative, the markers being *mai* (<*magis*) in Romanian and Megleno-Romanian, *cama* (*quam* + *magis*) in Aromanian, *më* in Albanian, and borrowed in Romani (generally the marker of the main contact language, but Slavic *po* and Turkish *da[h]a* are both more widespread). Remnants of a synthetic comparative in *-eder* also survive in some Romani dialects, but generally those spoken outside the Balkans. Given that Romani entered the Balkans some time between the 10th and 13th centuries, and given that during this same period Slavic preserved its inflectional system of adjectival gradation, it would appear that BS and Romani were undergoing this shift at about the same time, and those dialects that left the Balkans did so before its completion.

In general, the standard of comparison is an ablative marker, which is synthetic in Turkish and most of Romani but prepositional (lexical 'from') in BS, BR, Greek, some Romani, and Albanian, particularly Tosk. Albanian can also use relative *se* and BR can have relative *ca*. Clausal comparisons (e.g., 'to eat is better than to sleep') in Albanian, BR, and BS involve quantifiers, *se* [*sa*] 'that [how much]', *de.cît* 'from. how much', *ot.kolko[to]* 'from.how much [that]', respectively. Greek has *pará* 'contrary to, despite'.

There is a bifurcation in the superlative between Turkish and BS, on the one hand, and Greek and Albanian on the other, with BR and Romani occupying a middle ground. In Turkish and BS, the relative superlative is purely analytic and uses native markers: Turkish *en*, BS *naj*. In Greek and Albanian, the relative superlative is expressed by the definite of the comparative. (Greek also has a synthetic absolute comparative in a few adjectives.) Romanian and most of MR pattern like Albanian, whereas Aromanian and the MR of Tsárnareka have borrowed Slavic *naj*.

The expression of analytic adjectival gradation in Turkish is attested in the oldest monuments (8th century). The Greek dialects of Epirus, Thrace, Asia Minor, and of the Sarakatsan (transhumant Hellenophone shepherds) use the comparative marker [*a*]kóm[*a*] 'yet, still', calquing exactly Turkish *daha* (Table 1).

In Moldavian Gagauz, *sam* (<Russian *samyj*) is in competition with *en* as the superlative marker for the younger generation of speakers.

Table 1 Balkan adjectival gradation

| | | | | |
|-----------------|---------------|---------|----------|-------------------------|
| Turkish | daha | büyük | ablative | en büyük |
| Romani (Arlı) | po/da[h]a | baro | ablative | en/naj baro |
| Bulgarian | po- | goljam | ot | naj-goljam |
| Macedonian | po | golem | od | najgolem |
| Aromanian | kama | mari | di | nai mari |
| MR (Tsárnareka) | mai | mari | di | naimar[li]i most big |
| MR | mai | mari | di | tsāl mai mar[li]i |
| Romanian | mai | mare | de[cît] | cel mai mare |
| Albanian (Tosk) | më | i madh | nga | më i madhi |
| Greek | pio | megálos | apó | o pio megálos |
| | more | big | from | the more big |
| | 'bigger than' | | | 'biggest' |

Numeral Formation: The Teens The formation of teens by means of a construction meaning 'numeral on ten' is pan-Slavic but absent from Baltic, occurs in BR but not the rest of Romance, and is also Albanian. Although assumed to be a calque from BS into BR and Albanian, Hamp (1992) has pointed out that the words for 'twenty' in BS and BR and 'thirty' in Albanian show the numeral 'ten' is masculine in Slavic but feminine in Albanian and BR. Based on the isomorphism in gender for BR and Albanian and a combination of old shared sound changes and ancient borrowed lexicon among the three, Hamp suggests that this innovation occurred at a time when the Indo-European dialects that became Slavic, Albanian, and the language that Latinized into Romanian were part of a Northwest European sprachbund prior to their respective migrations to the Balkans (Table 2).

Analytic Subjunctive The analytic subjunctive formed by means of a subordinating particle (SP), usually of pronominal origin, plus a finite verb agreeing with its subject (omitted if the same as in the main clause, specified if different) replaces older nonfinite complements (infinitives) in all Balkan languages to varying degrees. Gheg has a new infinitive employing the preposition *me* 'with' and a short participle in contexts where Tosk uses the analytic subjunctive, but Gheg also has uses of the analytic subjunctive, and Tosk has some nonfinite participial constructions where other Balkan languages have the analytic subjunctive. Romanian and MR still have remnants of the Latin infinitive that can be used in some traditional infinitival functions. The BR infinitive is strongest in Maramureș, the northernmost Romanian region and the one in most contact with infinitive-using languages (Ukrainian, Hungarian, formerly Yiddish). BR in general also preserves Latin infinitives in *-re* as verbal nouns. Greek has a morphological remnant of the infinitive, but its only living function is to

Table 2 Balkan teens and tens^{a-f}

| | | | |
|-----------------|---------------------|------|----------|
| OCS | edinŭ | na | desęte |
| Romanian | un | spre | zeci |
| Aromanian | ună | spră | [dzătse] |
| MR | un | spră | ts |
| Albanian (Tosk) | një | mbë | dhjetë |
| | one | on | ten |
| Greek | enteka (en[a] deka) | | |
| | one ten | | |
| Romani | deš u jekh | | |
| | ten and one | | |
| Turkish | on bir | | |
| | ten one | | |
| | 'eleven' | | |

^aSlavic gender in numerals: dva (MASC) dve (FEM) 'two'.

^bRomanian gender in numerals: doi (MASC) două (FEM) 'two'.

^cAlbanian gender in numerals: tre (MASC) tri (FEM) 'three' (dy [MASC], dy [FEM] 'two').

^dOCS 10 = MASC dŭva desęte 'twenty'.

^eRomanian 10 = FEM două zeci 'twenty' (zece 'ten' < Lat. decem).

^fAlbanian 10 = FEM tri dhjetë 'thirty'.

represent the main verb in perfects and pluperfects. Bulgarian has a very marginal remnant of the Slavic infinitive limited to subordination to a tiny number of verbs. The infinitive has disappeared completely from Torlak except in some folk songs. Macedonian and Romani have eliminated all traces of earlier infinitives. Thus the replacement of infinitives with subjunctives is not uniform but scalar. At one end is Gheg, followed closely by Romanian, then Tosk, Bulgarian, Greek, and Vlah, with Torlak, Romani, and Macedonian at the other end.

New infinitival constructions have arisen in Romani outside of the Balkans in contact with infinitive-using languages. In Macedonian, some uses of the verbal noun can replace SP-clauses and thus function as a kind of new infinitive, although these constructions, which are highly colloquial, are merely alternatives. The option of using an SP-clause rather than an infinitive is available to all of BCS, but there is a

Table 3 Balkan SP clauses

| | | | |
|-----------------|-------------------|-----------|--------------------|
| Romani | mangav | te | hramonav |
| Albanian (Tosk) | dua | të | shkruaj |
| Albanian (Gheg) | [due | me | shkrue] |
| Greek | thélō | na | gráfo |
| Bulgarian | iskam | da | piša |
| Macedonian | sakam | da | pišuvam |
| Torlak | oču | da | pišem |
| Romanian | vreau | să | scriu |
| Vlah | voi | s(i) | scriu |
| <i>gloss</i> | <i>I.want</i> | <i>SP</i> | <i>I.write</i> |
| WRT | isterim | | yazayım |
| <i>gloss</i> | <i>I.want</i> | | <i>I.write.OPT</i> |
| | 'I want to write' | | |

tendency for such usage to become more frequent as one moves from northwest to southeast in the direction of Torlak. Since 1991, Croatian language planners have identified SP-clauses with Serbian and infinitives with Croatian, as a result of which Croatian speakers are now discouraged from using SP-clauses. In Serbian and Bosnian, however, the two constructions continue to coexist amicably (Table 3).

In WRT, optatives have expanded at the expense of infinitives owing to the influence of the other Balkan languages. The usage in Table 3 was a possibility in older Turkish, but, in a classic case of convergence via feature selection, the WRT optative now occurs where Turkish would normally have a nonfinite construction:

ben seni ist-er-im şimdi bir
 I you.ACC want-PRES-1SG now one
 müneccim ol-a-sın
 astrologer be-OPT-2SG
 Now I want you to be an astrologer

Similarly, Balkan Judezmo, which preserves the Spanish infinitive, nonetheless has some uses of its subjunctive, e.g., in questions, that calque Balkan SP-clauses and do not occur in Modern Spanish or North African Judezmo:

kwando ke te vengamoz a tom-ar?
 (Balkan Judezmo)
 when that you.ACC we.come to take-INF
 Pôte na 'rthoume na se pároume? (Greek)
 when SP we.come SP you.ACC we.take
 Koga da ti dojdeme da te zename?
 (Macedonian)
 when SP you. we.come SP you. we.take
 DAT ACC
 Cuando quieres que vengamos a recog-er-te?
 (Modern Spanish)
 when you.want that we.come to take-INF-you
 'When do you want us to come to get you?'

All Balkan languages use the independent analytic subjunctive to express wish, desire, or a milder form

of imperative. Albanian also has a synthetic optative used mostly in formulae.

Futures in 'Will' and 'Have' When Slavic entered the Balkans (6th–7th centuries C.E.), there was competition between the auxiliaries 'have' and 'want (will)' + infinitive to mark futurity in Latin and Greek, with Latin favoring 'have' and Greek favoring 'want'. OCS used the perfective of 'be' in addition to 'want', 'have' and various forms of 'begin' + infinitive. The 'will' + infinitive construction survives (with modified or new infinitives) in Romanian, northwestern Gheg (near and in Montenegro), in Bulgarian dialects (with postposed auxiliary), and MR (for speculations and threats). This form also survives in all the non-Balkan Štokavian dialects of BCS and connects them with East South Slavic. In fact, much of Štokavian ended up in its current location as a result of northward migrations during the 15th–18th centuries. The rest of Slavic developed the perfective of 'be' as a future marker. The next stage was 'will' + SP + conjugated present tense verb (Greek 14th century, Slavic 15th century). This stage also survives in BCS, including Torlak. The third stage, which overlaps the second, is the transformation of 'will' into an invariant particle + SP + conjugated main verb. This type of construction is still the main one in Tosk and parts of Gheg, especially in the northwest and southeast peripheries; it is characteristic of southern Romanian and survives in Torlak and in certain modal uses in East South Slavic, but not in Greek. The fourth stage is the elimination of the SP so that the future is marked by an invariant particle plus a conjugated verb. In addition to being the standard future in Balkan and southern Vlach Romani, Greek, and BS, it is common in colloquial Tosk. In MR, the future marker merged with SP, producing a new particle, *ās*, in Tsárnareka, but eliminating a distinct future marker in the other villages. Romani outside the Balkans has other means of forming or expressing the future, and it appears that the Romani development in the Balkans occurred in concert with the other Balkan languages (cf. 'Analytic Gradation of Adjectives') (Table 4).

Conjugated 'have' + infinitive, attested for early stages of all the traditional Balkan languages, remains the predominant future in most of Gheg. Conjugated have + SP + present is still used in Romanian, and invariant 'have' (which can also be an existential in all the Balkan languages with lexical 'have', cf. French *il y a*) is used in Arbëresh and occurs with modal functions in BS. In East South Slavic, the ordinary negated future uses this negative existential + SP + present, and this type is calqued into Aromanian, Romani, and WRT. Since Turkish and most of

Table 4 Balkan futures

| | | | |
|------------------------------|------------|--------|----------|
| Romani | ka | | dža[s] |
| Albanian (Tosk) | do | [tê] | shkojmë |
| Greek | tha | | páme |
| Bulgarian | šte | | trūgnem |
| Macedonian | ke | | odime |
| Torlak | če | | odime |
| Romanian (Colloquial, South) | o | să | mergem |
| Aromanian | va | s- | neadzim |
| MR | | si, să | neadzim |
| [MR-Tsārnareka | ās | | neadzim] |
| English | we will go | | |

Table 5 Negated futures

| | | | |
|------------|-------------|--------|-------------|
| Macedonian | nema | da | odime |
| Bulgarian | njama | da | hodime |
| Aromanian | noare | s' | neadzim |
| | not.has | SP | we.go |
| Romani | na-e | amen | te |
| | not-is | we.ACC | SP |
| WRT | yok-tur | | go.IPL.PRES |
| | not-is | | gid-elim |
| English | we won't go | | go-OPT.1PL |

Romani lack lexical verbs meaning 'have', their calques use their negated existentials, which also code possession (Table 5).

Future in Past as Conditional The combination of a future marker with a past tense marker to form a conditional, especially irrealis, is a classic Balkanism, although its realization differs among the various Balkan languages. (The construction itself can have a variety of related meanings, e.g., 'X almost happened/was about to happen', iterative-habitual, anterior future, and languages and dialects can be differentiated on the basis of which of these meanings are encoded.) Greek, Macedonian, and Romani all use the invariant future marker plus the imperfect. Tosk and Aromanian are almost the same, but they still have the SP, at least optionally. MR has an invariant 'will' marker (*vr̥ea*) + SP + present or perfect (see 'Perfect in "Have"'). In Bulgarian, Torlak, and other dialectal BCS, however, it is 'will' that conjugates in the imperfect + SP + present, and Gheg has the conjugated imperfect auxiliary 'have' + infinitive. The Balkan construction extends into BCS as far as southern Croatia and southwestern Serbia, and the southern Montenegrin dialects have the widest range of uses for the construction, thereby being most Balkan. In Turkish, the future participle plus a past auxiliary [*i]di* or [*i]miş* has the same nuances of irrealis conditional (Table 6).

Table 6 The Balkan conditionals

| | | | |
|---------------------|--------------------|---------------|-----------------------|
| Romani | ka | | keravas* |
| Greek | thá | | égrafa |
| Macedonian | ke | | napravev ^u |
| Aromanian | va | [s] | făceam ^u |
| Albanian (Tosk) | do | tê | bēja |
| | <i>FU</i> | <i>SP</i> | <i>do.IM.ISG</i> |
| MR | <i>vr̥ea</i> | si | am fat(ă) |
| | want.PRES.3SG | SU | do.PERF.1SG |
| Bulgarian | štjah | da | napravja |
| | ščaše/ščeše | da | napravim/ radim |
| | <i>want.3SG.IM</i> | <i>SU</i> | <i>do.PRES.1SG</i> |
| Albanian (Gheg) | [kishna | me | bâ] |
| | <i>I.have</i> | <i>with</i> | <i>do.PART</i> |
| Romanian | aș | fi | făcut |
| | <i>COND</i> | <i>be.INF</i> | <i>do.PAST.PART</i> |
| Turkish | yap | acak | tı m |
| | <i>ROOT</i> | <i>FU</i> | <i>PAST 1SG</i> |
| 'I would have done' | | | |

*Arli has a new imperfect formed by the long present + imperfect of '3SG/PLbe', e.g. *kerava sine*.

In Greek, Albanian, and Vlah, conditional constructions normally have a form of the 'will' morpheme. In BS, the Balkan conditional is in competition with the inherited conditional using the old optative of 'be' (invariant *bi* in Macedonian, conjugating in Bulgarian and Torlak) + old resultative participle. Romani dialects in contact with Slavic also use invariant *bi* + present as a conditional. In Romanian, a special conjugation of 'have' + infinitive serves as a conditional-optative.

Perfect in 'Have' The use of 'have' as an auxiliary with a nonfinite main verb to form an analytic perfect is attested for Greek and Latin at the end of the ancient period and is characteristic of Albanian, BR, and Greek, while such constructions (and lexical 'have') are absent from WRT and most of Romani. In BS 'have' + past passive participle (or its descendant) forms resultative constructions ranging from a fully grammaticalized perfect (with an invariant neuter verbal adjective) that has completely replaced the inherited perfect ('be' + old resultative particle in -/l) in extreme southwestern Macedonian and spreading north to Mt. Šar and east to the Vardar and beyond, to resultative syntagms with 'have' + past passive participles agreeing with their direct objects and limited to transitive verbs with human subjects in most of Bulgaria.

Given the geography and history of the 'have' perfect in BS, it is clearly a calque on one of the non-Slavic contact languages. Although Greek and Albanian have been proposed as the possible models, Gołąb's arguments in favor of Aromanian are the most convincing. In Aromanian, the feminine participle is selected

as the invariant, since in BR (as in Albanian) the feminine gender is unmarked (neuter is obsolete). The Macedonian invariant neuter verbal adjective therefore corresponds exactly to the Aromanian in terms of unmarked gender. In Greek, the main verb is a remnant of the infinitive and in Albanian the participle does not mark singular gender. Thus the BR construction most closely resembles the Macedonian. An additional argument in favor of BR as the model is evidence of Macedonian and Vlah mutual calquing in other resultative constructions.

Evidential In a Balkan context, evidentiality (inferential, distance, mode of indirect narration, indirective, status, French *médiatif*) is a grammatical category encoding the speaker's evaluation of the narrated event, often, but not always, predicated upon the nature of the available evidence. Evidentials can be of two types: confirmative (vouched for, 'witnessed') and nonconfirmative (not vouched for, 'reported', 'inferential'). The nonconfirmative can be felicitous (neutral report or inference) or infelicitous, in which latter case the nonconfirmative expresses either acceptance of a previously unexpected state of affairs (i.e., surprise, admirativity *sensu stricto*) or rejection of a previous statement (i.e., sarcasm, dubitativity). The opposition confirmative/nonconfirmative was already encoded in the Turkic simple past in *-di* (confirmative) and the perfect participle in *-miş* (nonconfirmative) at the time of the earliest monuments. In East South Slavic, the old synthetic pasts are markedly confirmative (this same meaning is also sometimes identified in Torlak). By contrast, the old perfect using the resultative participle in *-l* has become an unmarked past, with a chief contextual variant meaning of nonconfirmative. In Albanian, the inverted perfect (participle + 'have') has fused into a marked nonconfirmative present paradigm called admirative, which can then function as an auxiliary to form analytic past tenses. The Frasheriote Aromanian dialect of Bela di Suprā has reinterpreted the 3SG.PRES Albanian admirative marker as an admirative suffix, which it adds to a masculine plural

imperfect participial base to form a new admirative (Table 7).

Megleno-Romanian uses an inverted perfect + auxiliary construction in a similar function. The Romanian presumptive mood formed with a future, subjunctive, or conditional marker + invariant *fi* 'be' + gerund (or past participle) is a similar marked nonconfirmative, as is the probabilitive mood (based on a BCS-type inverted future) of Novo Selo Bulgarian, a dialect spoken across the Danube from Romania and a few kilometers east of Serbia (Table 8).

The Judezmo of Istanbul uses the pluperfect as a calque on the Turkish past in *-miş*:

Kuando esta-v-an en l' Amérika, les
 when be-IM-3PL in the America them.DAT
 av-iy-a entra-do ladrón
 have-IM-3SG enter-PAST.PART thief
 'When they were in America [i.e., absent], a thief broke into
 (Turkish *girmiş*) their house.'

Other Many other features too numerous to discuss here are cited as Balkanisms, e.g., the conflation of adverbs of location and motion ('where'/'whither'), purposives in 'for' + SP + verb and other prepositional parallelisms, a distinction between realis and irrealis complementizers, and absolute relativizers and interrogatives as complementizers, this last being a feature that has spread to WRT:

čovек-ot što go vid-ov (Macedonian)
 person-the what him.ACC see-1SG.AOR
 adam ne cür-d-üm (WRT)
 man what saw-PAST-I
 gör-düğ-üm adam (Standard Turkish)
 see-PART-my man
 'the man that I saw'

Word Order

Clitic Ordering Greek, Albanian, and BR all permit absolute initial pronominal clitics when the first stressed element is a finite verb, but in BS only Macedonian (especially the western dialects) permits this. Bulgarian keeps pronominal clitics bound to the verb but either requires the verb or some other element in

Table 7 Aromanian (Färshälots, Bela di Suprā and Albanian indicatives (3sg 'work')

| | Nonadmirative | | Admirative | |
|-----------------------|------------------|------------------|--------------------|---------------------|
| <i>Present</i> | lukrā | punon | lukracka | punuaka |
| <i>Perfect</i> | ari lukratā | ka punuar | avuska lukratā | paska punuar |
| <i>Pluperfect</i> | ave lukratā | kish punuar | — | paskësh punuar |
| <i>2nd Pluperfect</i> | avu lukratā | pat punuar | — | — |
| <i>Double perfect</i> | ari avut lukratā | ka pasë punuar | ari avuska lukratā | paska pasë punuar |
| <i>Double plup.</i> | ave avut lukratā | kish pasë punuar | ave avuska lukratā | paskësh pasë punuar |
| <i>2nd Dbl. plup.</i> | avu avut lukratā | pat pasë punuar | — | — |

Table 8 The Novo Selo probabilitive 'see'

| | | | |
|---------|---|---------------------------------|-------------|
| Present | 1 | gla*dăcă m | gla*dăcă mo |
| | 2 | gla*dăcăș | gla*dăcă tă |
| | 3 | gla*dăcă | gla*dăcă ju |
| Future | | čă gla*dăcă m, etc. | |
| Past | | budăcă m ~ bičă m glă dâl, etc. | |

initial position. BCS, including most of Torlak, still follows Wackernagel's law and has clitics in second position.

Az često mu go davam. (Bulgarian)
 I often. him.DAT it.ACC give.1SG.PRES
 Ja mu ga često dajem. (Serbian)
 I him.DAT it.ACC often give.1SG.PRES.
 'I often give it to him'
 Davam mu go (Bulgarian)
 give.1SG.PRES him.DAT it.ACC
 Mu go davam (Macedonian)
 him.DAT it.ACC give.1SG.PRES
 'I give it to him'

Romani pronominal clitics follow the verb. WRT is basically suffixal, like the rest of Turkish, and clitics always follow the stressed element, but elements that can be fused or separated are more likely to be separated and less likely to show vowel harmony in WRT.

Constituent Order Balkan languages are characterized by relatively free constituent order with certain patterns being favored for various types of syntactic and narrative strategies (emphasis, topicalization, focus, contrastive thematization, etc.). The unmarked word order tendency is SVO in all the Indo-European Balkan languages. Unlike most of Turkish, where the tendency is verb-final, WRT and Gagauz show SVO tendencies. Similarly, BR, BS, Albanian, and Greek all have the basic order head-genitive, while Turkish and Romani are genitive-head. Romani dialects in the Balkans and WRT, however, also have head-genitive constructions:

m-e phral-es-k(er)e kher-es-k(or)o vudar (Romani)
 my-OBL brother-OBL- house-OBL- door
 GEN GEN
 'the door of my brother's house'
 o vudar e kher-es-ko
 the.MASC.NOM door the.OBL house-OBL-GEN
 m-e phral-es-kere (Romani)
 my-OBL brother-OBL-GEN
 Baba-si Ali-nin (WRT)
 father-his A.-GEN
 Tathto mu na Ali (Macedonian)
 father him.DAT to Ali
 Baba-i i Ali-ut (Albanian)
 father-DEF PC.MASC.NOM.SG A.-Def.GEN
 Ali-nin babası (Standard Turkish)
 Ali-GEN father-his
 'Ali's father'

Adjectives generally follow their heads in Albanian and BR, but precede in BS, Greek, WRT, and Romani. In all of these languages, the opposite order is possible in various discourse functions. Albanian enclaves in the eastern Balkans also have preposed adjective as the standard order.

Lexicon, Semantics, and Derivational Morphology

The etymological commonalities of the Balkan lexicon received considerable attention during the formative years of Balkan linguistics, whereas more recently the focus has been on shared grammatical features. Miklosich's 1861 survey of Balkan grammatical commonalities occupied only 4% of what was basically a study of the Slavic lexical influence on Romanian. Sandfeld (1930) devotes 40% of his book to the lexicon, whereas Asenova (2002) allots 10% of her text to such issues. Although the lexicon is the most salient surface manifestation of linguistic influence, words can travel between languages without the aid of communal multilingualism, whereas the diffusion or convergence of grammatical structures is a more complex process that requires at least a core community of bi- or multilingual speakers. In terms of the definition of a sprachbund, it is the shared grammatical features rather than shared vocabulary that is the key determiner, although shared vocabulary is usually part of the picture.

There are common loanwords from each of the component language families in the Balkan languages. Words shared by Albanian and Romanian of pre-Latin (substrate) origin are often connected with domestic items or husbandry, e.g., Albanian *shtrungë*, BR *strungă*, BS (Macedonian and west Bulgarian) *strunga*, Greek (Epirus and Sarakatsan) *stroúgka* 'dairy'. Greek, Slavic, and Romance (especially Balkan Latin and Venetian Italian) were all languages of power in the Balkans at various times during the Middle Ages and contributed a variety of lexemes and even derivational affixes to the common Balkan lexicon, e.g., the Latin agentive suffix *-arius*, the Slavic feminine suffix *-ica*, and the Greek aorist marker *-s-* (used in deriving verbs). As the language of administration, the market place, and urban life in general, Turkish dominated the Balkan peninsula for more than half a millennium. By the 19th century, the shared Turkish lexicon in the Balkan languages was of considerable size. The rise of Balkan standard languages, however, entailed the stylistic lowering and marginalization of many Turkish loanwords, and as many of these items were of Arabo-Persian origin, they were discouraged by Turkish purists as well. The Turkish agentive *-ci*, attributive *-li*, qualitative or concrete *-lik* (with adjustments for vowel harmony, voicing

assimilation, and adaptation) continue to be productive as derivational affixes, e.g., Macedonian *puberteti* 'adolescent (ironic)' Albanian *partiakçi* 'party hack', Judezmo *hanukalik* 'Chanukah present', etc.

The Balkan languages also share numerous idioms, collocations, and calqued expressions; e.g., the use of 'eat' to mean 'undergo something unpleasant' as in 'eat wood' = 'take a beating' or 'it doesn't cut his mind' = 'he doesn't understand'. There are a variety of shared discourse particles and conjunctions (e.g., Turkish *am[m]a* 'but', Greek *bre* 'hey, vocative particle') that also form part of the common Balkan lexicon.

Sociolinguistics

Factors such as power, prestige and religion have influenced directions and degrees of Balkan contact phenomena. Throughout the Ottoman period, Turkish had high prestige as the language of the state and the town, Greek had prestige among Christians as the language of the (Orthodox) church with its own literary tradition (and history of power, i.e., Byzantium) and was also a language of commerce. BS had less prestige in the southern Balkans, but its history of medieval literacy and political competition with Byzantium gave it some limited prestige. Although BR was descended from Latin, another language of empire and conquest, the local varieties that developed after the Slavic invasions did not have that level of prestige and, like Albanian, were associated mainly with rural contexts. In Wallachia and Moldavia, Church Slavonic was the liturgical language for centuries, and Romanian was written in Cyrillic until the mid-19th century. Aromanian speakers in southern Balkan towns used Greek outside the home. Romani was at the bottom of the social hierarchy, but Judezmo was outside it. This is reflected in 19th-century Macedonian folklore collections, where characters in ethnic jokes, including Roms (Gypsies), speak in their own languages, except Jews, who speak Turkish, not Judezmo. For both Romani and Judezmo, multilingualism was unidirectional, i.e., Roms and Jews learned other languages but heard their languages spoken by others rarely, if ever. At the opposite end of the prestige scale, speakers of Greek and Turkish were less likely to learn less prestigious languages but were more likely to hear their languages spoken by others. Those languages in the middle of the hierarchy (BS, BR, and Albanian) had the highest degree of multidirectional multilingualism and show a higher degree of congruence.

Marriages could be freely contracted across linguistic lines but not religious ones, so that multilingual households were a commonplace. Although speakers

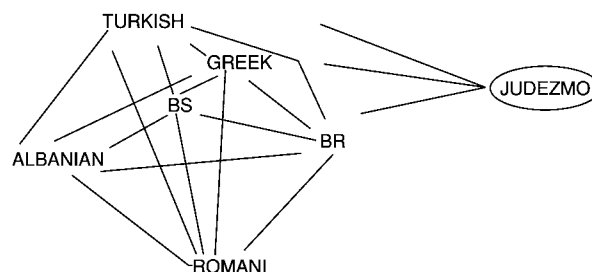


Figure 1 Schematic linguistic social/political hierarchy (Ottoman Period).

of BS, BR, and Greek were mostly Christian and speakers of Albanian were usually Muslim, each of these religions also had significant communities speaking the other languages. Except for Gagauz, speakers of Turkish were Muslim, but there was still plenty of linguistic contact via religious conversion. Jews and Roms, however, were endogamous along a combination of linguistic and other social lines. This boundary maintenance is reflected linguistically in Romani, where there is a clear opposition between the relatively open systems of adjectival comparison and modality on the one hand to the conservative nominal, pronominal, and tense-aspect systems on the other.

Figure 1 illustrates the relative prestige of the various languages during the Ottoman period. Height symbolizes prestige, while incline indicates relative (never absolute) directionality. The directionality is reversed in the case of slang and secret languages, where it is the covert prestige of languages further down on the social scale that is reflected in patterns of lexical borrowing. In the case of Judezmo, knowledge of Turkish was most widespread, while knowledge of other Balkan languages would depend on the particular (urban) environment.

Causation

For most of the history of Balkan linguistics, causation has been sought in the influence of (interference from) one of the languages, e.g., Greek, Latin, or a pre-Latin non-Hellenic substratum (e.g., Illyrian, Thracian, and/or Dacian – all so poorly attested that we do not have so much as a single sentence in any of them). More recently, however, an ecological model of feature selection argues that those grammatical developments more suitable for effective communication that might be already present in the language, i.e., more adaptive, are more likely to be selected for further development and spread (cf. 'Resumptive Clitic Pronouns [Reduplication, Replication]'). In such a model, languages can utilize native resources that are reinforced by their occurrence, or potential

for occurrence, in the contact languages. Mechanisms such as fusion, metatypy, and code copying are all potentially relevant. At the same time, sociolinguistic factors such as those adduced in 'Sociolinguistics' can influence directions of change. The diffusion of borrowings and the development of convergences are thus compatible parts of a larger picture of a sprachbund in which languages come to be similar without becoming identical. It is worth emphasizing here the insight of Joseph (2001), namely that the move from lexical via phraseological to syntactic borrowings that characterizes the contact-induced changes of a sprachbund such as the Balkans are quintessentially surface phenomena.

Although some scholars have argued against the idea of a Balkan sprachbund since the 1930s, the argument that the Balkans are basically just part of a larger European linguistic zone coincides roughly with the recent rise in interest in contact linguistics and typology. In the case of the Balkans, however, while it is clear that Kopitar's formulation is an exaggeration, it is equally clear that Trubetzkoy's original insight captures facts about language relationships. Of particular significance is the manner in which patterns map such that the languages that surround the Balkan sprachbund do not share the most salient features. The fact that English and Western Romance have gone even further than most of the Balkan languages in some changes does not contradict the hypothesis that the Balkan sprachbund is precisely that, i.e., a product of the process of language contact. If some of those contact-induced changes are the result of shared feature selections, having parallels elsewhere, that may contribute to identifying likely directions of language change, but it does not vitiate the sprachbund as a historical and sociolinguistic phenomenon.

In a sense, a sprachbund is more like a dialect chain than a linguistic family: as features spread over areas, they may do so with differential impact. Thus, while it is possible to define a sprachbund in terms of languages displaying a coalescence of a number of such features, it is not necessarily the case of an 'all and only' phenomenon. Moreover, the transition from pragmatic to syntactic (grammaticalized) to morphological sometimes maps onto the territory of the sprachbund itself, moving from periphery to core. Like dialects, there can be a transitional effect, and a given language, e.g., BCS, can participate in the changes to a greater or a lesser extent. For both the dialect and the sprachbund, politics can have a crucial effect in setting boundaries that favor internal consistency and external differentiation. Just as the very concept of language vis-à-vis dialect (e.g., to which language a given dialect 'belongs' or which isoglosses

will be chosen as defining one dialect in opposition to another) can be a complex of intersecting factors, so too can the definition of sprachbund.

Bibliography

- Alexander R (2000). 'Tracking sprachbund boundaries: word order in the Balkans.' In Gilbers D, Nerbonne J & Schaeken J (eds.) *Studies in Slavic and General Linguistics 28: Languages in contact*. Amsterdam: Rodopi. 9–27.
- Asenova P (2002). *Balkansko ezikoznanie. Osnovni problemi na balkanskija ezikov sajuz* (Balkan linguistics: Basic problems of the Balkan linguistic league) (2nd edn.). Sofia: Farber. (Resumé in English, French, and Russian).
- Atanasov P (1990). *Le mégléno-roumain de nos jours*. (Balkan-Archiv Neue Folge – Beiheft 8). Hamburg: Buske.
- Belyavski-Frank M (2003). *The Balkan conditional in South Slavic*. Munich: Otto Sagner.
- Friedman V (2003a). 'Evidentiality in the Balkans.' In Aikhenvald A & Dixon R (eds.) *Studies in evidentiality*. Amsterdam: Benjamins. 189–218.
- Friedman V (2003b). *Turkish in Macedonia and beyond: studies in contact, typology, and other phenomena in the Balkans and the Caucasus*. Wiesbaden: Harrassowitz.
- Friedman V (2004). *Studies on Albanian and other Balkan languages*. Peja: Dukagjini.
- Friedman V & Joseph B (2006). *The Balkan languages*. Cambridge: Cambridge University.
- Gołąb Z (1984). *The Arumanian dialect of Kruševo in SR Macedonia SFR Yugoslavia*. Skopje: Macedonian Academy of Arts and Sciences.
- Gołąb Z (1997). 'The ethnic background and internal linguistic mechanism of the so-called Balkanization of Macedonian.' *Balkanistica* 10, 13–19.
- Gălăbov I, Georgiev V & Zaimov J (eds.) (1968). *Actes du premier congrès international des études balkaniques et sud-est européennes* (vol. 6). Sofia: Bălgarska Akademija na Naukite.
- Hamp E (1982). 'The oldest Albanian syntagma.' *Balkansko Ezikoznanie* 25, 77–79.
- Hamp E (1992). 'Albanian.' In Gvozdanović J (ed.) *Indo-European numerals*. Berlin: Mouton de Gruyter. 835–921.
- Hinrichs U (ed.) (1999). *Handbuch der Südosteuropa-Linguistik*. Wiesbaden: Harrassowitz.
- Ilievski P (1988). *Balkanološki lingvistički studii* (Balkanological linguistic studies). Skopje: Institut za makedonski jazik.
- Joseph B (1983). *The synchrony and diachrony of the Balkan infinitive: a study in areal, general, and historical linguistics*. Cambridge: Cambridge University Press.
- Joseph B (1987). 'A fresh look at the Balkan Sprachbund: some observations on H. W. Schaller's *Die Balkansprachen*.' *Mediterranean Language Review* 3, 105–114.
- Joseph B (2001). 'Is Balkan comparative syntax possible?' In Rivero M & Ralli A (eds.) *Comparative syntax of Balkan languages*. Oxford: Oxford University Press. 17–43.

- Katičić R (1976). *Trends in linguistics, state-of-the-art report 4, 5: ancient languages of the Balkans* (2 vols). The Hague: Mouton.
- Koneski B (1967). *Gramatika na makedonskiot literaturni jazik* (Grammar of the Macedonian literary language). Skopje: Kultura.
- Lindstedt J (2000). 'Linguistic Balkanization: contact-induced change by mutual reinforcement.' In Gilbers D, Nerbonne J & Schaeken J (eds.) *Studies in Slavic and General Linguistics 28: Languages in contact*. Amsterdam: Rodopi. 231–246.
- Matras Y (1994). 'Structural Balkanisms in Romani.' In Reiter N (ed.) *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa*. Wiesbaden: Harrassowitz. 195–210.
- Nikolaeva T (1996). *Prosodija Balkan: Slovo – vyskazyvanija – tekst* (Prosody of the Balkans: Word – Utterance – Text). Moscow, Nauka.
- Reiter N (ed.) (1983). *Ziele und Wege der Balkanlinguistik* (Balkanologische Veröffentlichungen, Vol. 8). Berlin: Osteuropa Institut an der Freien Universität Berlin.
- Sandfeld K (1930). *Linguistique balkanique*. Paris: Klincksieck. (1926. *Balkanfilologien*. Copenhagen: Luno.)
- Sawicka I (1997). *The Balkan Sprachbund in light of phonetic features*. Toruń: Energeia.
- Schaller H (1977). *Bibliographie zur Balkanphilologie*. Heidelberg: Carl Winter.
- Sobolev A (2005). *Malyj dialektologičeskij atlas balkanskih jazykov*. (Small dialectological atlas of the Balkan languages). Marburg: Biblion Verlag.
- Solta G (1980). *Einführung in die Balkanlinguistik mit besonderer Berücksichtigung des Substrats und des Balkanlateinischen*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Topolińska Z (1995). 'Convergent evolution, creolization and referentiality.' In Hajičová E *et al.* (eds.) *Prague Linguistic Circle Papers*, vol. I. Amsterdam: Benjamins. 239–247.
- Varol M-C (2001). 'Calques morphosyntaxiques du turc en judéo-espagnol: mécanismes et limites.' In Donabedian A (ed.) *Faits de langues: langues de diaspora*. Paris: Ophrys. 85–99.

Balochi

P O Skjærø, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Balochi (or Baluchi, in several dialects) is spoken by the Baloch in eastern Iran and western Pakistan (Baluchistan), but also in southern Afghanistan, Turkmenistan, and the Arab Gulf States (totaling 6–8 million speakers?). The Baloch are first mentioned in literature about 1000 C.E., but the language did not become a written one until the 20th century, although the earliest known manuscript dates from the early 19th century. On the other hand, the Baloch have an oral poetic tradition with historical themes reaching back to the 15th century, but especially productive in the 19th century. Modern literature and publications are centered in Quetta in Pakistani Baluchistan and in Karachi. A Balochi Academy was founded in Quetta in 1959 and still publishes Balochi literature and supports Balochi language and culture in various ways, and the University of Quetta offers a Balochi Studies program. Balochi radio programs are broadcast from Zahedan in Iranian Baluchistan and from Quetta and Karachi, formerly also from Kabul.

There are, by one count, six principal dialects of Balochi, characterized by differences in grammar and

lexicon. The western dialect of Raxšāni is the largest, the principal subdialect being Sarhaddi.

Balochi belongs with the North(west) Iranian languages, differently from Persian, which is a South-west Iranian language; compare, for instance, Balochi *asin* 'iron,' *jan-* [dʒan-] 'strike,' *zird* 'heart,' versus Persian *āhan*, *zan-*, *dil*. It is a phonetically conservative language, having preserved much of the Old Iranian consonant system intact, notably intervocalic stops and affricates, for instance, Bal. *pād* 'foot,' *āp* 'water,' *roč* [rōtʃ] 'day' (Pers. *pā*, *āb*, *rūz*). Among innovations are the development of initial *w-* to *g(w)* (OIran. *wāta-* 'wind,' Bal. *gwāt*, Pers. *bād*), *xw-* to *w-* (OIran. *xwara-* 'eat,' Bal. *war-*, Pers. *x^oor-*), and the change of fricatives into stops (Bal. *nākun* 'nail,' Pers. *nāxon*; Bal. *gipta* 'seized,' Pers. *gereft*).

Balochi has retroflex consonants in words borrowed from Indo-Aryan, including originally English words, for instance, *ḍrēwar* (*ḍ* = [ɖ]) 'driver.'

There is a four-case system, distinguishing nominative, genitive, and an oblique case. The suffix *-rā* (*-ā* with personal pronouns) can be added to the oblique to express direct and indirect objects.

Notable features of the verb system include the formation of continuous tenses by means of a present participle in *-ag* (*raw-ag-ā int* 'go-ing-in he-is' = 'he is going'), a construction perhaps influenced by

the neighboring Indic languages and replacing the older formation with prefix *a-* (*a-rot* 'he goes, is going'). In the conservative dialects, the past tenses use the ergative construction in pure passive form (*man gūnī zurt-ant o šut-un* 'I.OBL sack.PL take.PAST-3RD.PL and go.PAST-1ST SING' = 'I took the sacks and went'), while in the western dialects, the active construction of Persian prevails (*man gūnī-ān zurt-un o šut-un* 'I.DIR/OBL sack.PL.DO take.PAST-1ST.SING and go.PAST-1ST.SING' = 'I sacks took-I and went-I').

Balochi lexicon contains a large number of loanwords, mainly Arabo-Persian and Indo-Aryan from a western Sindhi dialect, as well as a small number of words from Brahui, a Dravidian language, which contains a large number of Balochi words.

Bibliography

- Badalkhan S (1994). *Poesia epica e tradizioni orali balochi. I menestrelli Pahlawân del Makran*. Ph.D. diss., University of Naples.
- Elfenbein J (1989a). 'Balōči.' In Schmitt R (ed.) *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert. 350–362.
- Elfenbein J (1989b). 'Baluchistan iii. Baluchi language and literature.' In Yarshater E (ed.) *Encyclopaedia Iranica*, vol. II. London: Routledge and Kegan Paul. 633–644.
- Korn A & Jahani C (eds.) (2004). *The Baloch and their neighbours. Ethnic and linguistic contact in Balochistan in historical and modern times*. Wiesbaden: Reichert.
- Mills M A, Claus P J & Diamond S (eds.) (2003). *South Asian folklore. An encyclopedia. Afghanistan, Bangladesh, Nepal, Pakistan, Sri Lanka*. New York: Routledge.

Balto-Slavic Languages

S Young, University of Maryland Baltimore County, Baltimore, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

The term Balto-Slavic encompasses the languages of the closely related Baltic and Slavic branches of the Indo-European language family. The Slavic languages, traditionally divided into East, West, and South Slavic, are well-represented over much of Central and Eastern Europe and Siberia. Of the once more numerous and widespread Baltic languages, only two have survived to the present, Latvian and Lithuanian, which together form East Baltic. West Baltic is represented by Old Prussian, which died out in the early 18th century; it is known from word lists, place names, and catechism translations.

The nature of the relationship between the Baltic and Slavic languages has long been a source of debate. In the traditional *Stammbaum* approach, reflected in K. Brugmann's landmark *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen*, Baltic and Slavic are presented as equivalent branches of a Balto-Slavic protolanguage, which derives in turn from Proto-Indo-European.

The assumption of a post-Indo-European period of Balto-Slavic linguistic unity is based on a number of striking and seemingly exclusive correspondences between the Baltic and Slavic languages. In phonology, the most cogent argument for a Balto-Slavic protolanguage is found in the highly complex prosodic structures of both language families, which typically

agree in details of stress placement (including reflexes of Hirt's law), syllable tone (including reflexes of Winter's law), and accentual paradigm. Among other phonological agreements, syllabic resonants develop into both *-iR-* and (exceptionally) *-uR-*, with a similar distribution in both language groups. Morphological correspondences include an *-ā* formant marking the preterite/aoist stem, often accompanied by the reduced grade of the root: Lith. *pīrko* (pres. *peřka*) 'buy.PRET' : OCS *žbda* (pres. *židetŭ*) 'wait, expect.AOR'; and a present passive participle in *-m-* (for East Baltic and Slavic): Lith. *nešamas*, Latv. *nesams* : OCS *nesomŭ* 'being carried'. There are a number of exclusive correspondences in word formation, among them deverbal nouns in *-imo-*: Lith. *piešimas* 'drawing' : Slavic **pismo* 'writing'; agent nouns in *-ā-jo-*: Lith. *artójus*, OPr. *artoys* : OCS *ratajb* 'plowman'; agent nouns in *-ik-o*: Lith. *siuwikas*, OPr. *schuwikis* 'shoemaker' : ORus. *šuvьсь* 'tailor; shoemaker'; denominal adjectives in *-in-*: Lith. *krūvinas*: OSC *krūvьnъ* 'bloody', and diminutives in *-uk-o-*: Latv. *dēluks*, OCS *synŭkŭ* 'sonny'. Finally, there are many apparently exclusive lexical items, among them Lith. *rankà* : OCS *rŭka* 'hand'; Lith. *rāgas* : OCS *rogŭ* 'horn'; Lith. *liepa* : OCS *lipa* 'linden'.

The assumption of a Balto-Slavic proto-language was first challenged by A. Meillet (1908), who argued that the various agreements between the two language families are only apparent, a result of inherited archaisms and parallel developments in each of the branches. A refinement of this model was advanced by J. Endzelin (1911), who accounted for shared features

by positing a period of prolonged language contact between neighboring Baltic and Slavic communities, leading to a degree of linguistic convergence.

More recent studies have stressed the non-equivalence of the notions of Baltic and Slavic. C. Stang (1966: 10 ff.), developing one of Endzelin's ideas, pointed out that while Common Slavic presents a relatively uniform system, Baltic is divided by a number of significant isoglosses. Certain of the innovations represented by these isoglosses connect East Baltic with Slavic as opposed to Old Prussian (for example, the Indo-European **-s(i)o* genitive singular of *o*-stem nouns, apparently preserved in Old Prussian, has been replaced in East Baltic and in Slavic by a form that appears outside of Balto-Slavic in ablative function: Lith. *rāgo*, Latv. *raga* = OCS *roga*, 'horn.-GEN SG'); while other isoglosses link Old Prussian with Slavic (for example, the possessive pronouns OPr. *mais*, *twais*, *swais* = OCS *mojъ*, *tvojъ*, *svojъ* 'my, your, one's own' are refashionings of the IE root represented in Lith. (*manas*), *tavas*, *savas* and Latvian (*mans*), *tavs*, *savs*).

V. V. Ivanov and V. N. Toporov (1961), in reviewing the methodological preconditions for discussing the Balto-Slavic relationship, have argued that a relatively homogeneous proto-Slavic can be derived from a considerably more archaic and heterogeneous proto-Baltic linguistic model, in effect redefining the notion of Balto-Slavic by treating Slavic as a local development within a Baltic dialectal continuum.

Progress in further defining the relationship between Baltic and Slavic is hampered by a lack of linguistic data from the former Baltic populations assimilated by the East Slavs in the upper Dniepr river basin (the Dniepr Balts), and in present-day

Baltic territories by neighboring Latvians and Lithuanians (the Couronians, Selians, Zemgalians, and Jatvingians). The written documents of extinct Old Prussian are scant and rather unreliable, while the earliest monuments of Latvian and Lithuanian date only from the 16th century, when these languages already had a modern appearance. Nevertheless, dialectal data (including toponymic) still being drawn from various Baltic and Slavic languages, together with a more profound study of Baltic and Slavic borrowings in neighboring languages, may help provide new perspectives on the question of Balto-Slavic.

Bibliography

- Birnbaum H (1970). 'Four approaches to Balto-Slavic.' In Rūke-Dravina V (ed.) *Donum Balticum*. Stockholm: Almqvist & Wiksell. 69–76.
- Brugmann K (1897). *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen*. Strassburg: K. J. Trübner.
- Dini P (2000). *Baltų kalbos: lyginamoji istorija*. Vilnius: Mokslo ir enciklopedijų leidybos institutas.
- Endzelin J (1911). *Slavjano-baltijskie ètjudy*. M. Zil'berg: Kharkov.
- Ivanov V V & Toporov V N (1961). 'K postanovke voprosa o drevnejšix otnošenijax baltijskix i slavjanskix jazykov.' In Tolstoj N (ed.) *Issledovanija po slavjanskomu jazykoznaniju* (4th International Congress of Slavists, Moscow, 1958). Izdatel'stvo Akademija nauk SSSR: Moscow. 273–305.
- Meillet A (1908). *Les dialectes indo-européennes*. H. Champion: Paris.
- Stang C (1966). *Vergleichende Grammatik der Baltischen Sprachen*. Oslo: Universitetsforlaget.

Bantu Languages

D Nurse, Memorial University of Newfoundland,
St. John's, NL, Canada

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Bantu is the largest of the dozen or so language families that make up the Niger-Congo phylum, which, with nearly 1500 languages, is the largest phylum in the world. Some 750 million people live in Africa, some 400 million speak a Niger-Congo language, and some 250 million – a third of all Africans – speak a Bantu language. Bantu-speaking communities live south of a

line from western Cameroon across the Central African Republic, the Democratic Republic of Congo (DRC; once known as Zaire), Uganda, and northern Kenya to southern Somalia. Most languages spoken from that line to the southern tip of South Africa are Bantu. Within that area, they coexist with some non-Bantu languages: a few Khoisan, mostly in the southwest, a few Cushitic, in the northeast, and a string of Nilo-Saharan and Adamawa-Ubangian languages along and within the northern border. In all, 27 African countries – roughly a half – are partly or entirely Bantu speaking.

Certain generalities are true of Bantu-speaking communities. One is that, just as most African

countries are multilingual, many individuals are bi- or multilingual. In former times, many who came in contact with neighboring communities, mainly males such as traders or soldiers, spoke two or more languages. Though this is still true, it is increasingly true that many young people are born into one language community, are formally educated in a second language, and may acquire a third language later. Men tend to speak more languages than do women, and living in cities encourages multilingualism more so than does life in rural areas (Wolff, 2000).

Another truism is that many languages are poorly described. Three broad language categories can be distinguished. At the lower end are many dozens of languages for which no data are available. In the middle, the largest set, are many languages for which some data are publicly available, ranging from just a word list to a partial description. Finally, for about 10% of the languages, a reasonably comprehensive grammar exists, as books, doctoral theses, or long articles. In some areas – for example, South and East Africa – the languages are fairly well described, whereas in others (Angola, Cameroon, the DRC, and Zambia) they are less well covered.

A third point concerns the poor health of some of these communities. Dividing the Bantu population of 250 million by the number of languages, 500, gives an average of half a million speakers per community, but that ignores the fact that many smaller communities, especially rural, are in much worse shape than the figures indicate. The best demographic collection available (Gordon, 2004) gives figures for most communities, but gives no breakdown according to age. In many small communities, the fluent speakers are aging, with few or no younger speakers, so these communities of speakers will silently fade away in this century. As the small communities get smaller, the large get larger. At the same time, new urban and regional forms of languages are thriving (see Sommer, 1992; Bernstein, 1998; Wolff, 2000).

Finally, across the area, it has proved difficult to distinguish language from dialect. The difference is one of degree of similarity, and the question concerns where to cut a cline of similarity. Thus readers should treat with skepticism the figure of 500 Bantu languages. Estimates have varied between 300 and 600. If lack of reasonable mutual intelligibility with other varieties is a major defining feature of a language, then the figure is nearer 250 than 500.

Classifications

The second half of the 20th century saw dozens of referential and genealogical classifications. The most widely used referential system is that of Guthrie

(1948, 1971), who divided the (Narrow) Bantu languages into 15 zones (designated A, B, C, D, E, F, G, H, K, L, M, N, P, R, S), and each zone in turn into groups (designated A11, A12, A13, ..., etc.), for a total of 85 groups. Thus Nen is A44, Lingala is C36(d), Ha is D66, varieties of Swahili are G41, -42, or -43, and Zulu is S42. Guthrie's zones and groups are based partly on shared features he regarded as important, and partly on geographical contiguity. The most recent version of Guthrie (2003) is Maho. Guthrie's taxonomy did not reflect history, except indirectly.

In contrast, genealogical classifications aim to reflect the evolutionary history and, to a lesser extent, the contact history of the Bantu languages. Nearly all genealogical classifications assume that the current languages derive from Proto-Bantu and nearly all are based on the use of vocabulary in some form: lexicostatistics (counting percentages of shared vocabulary), glottochronology (assigning dates to percentages of shared vocabulary), shared lexical innovations, or juxtaposing results from lexical investigations with those from other disciplines, such as culture or archaeology (DNA comparison on a wide scale is so far lacking). The only study that examined the whole Bantu area and drew on hundreds of languages was the lexicostatistical work of Bastin *et al.* (1999); others used a smaller sample (Ehret, 1998, 1999; Nurse, 1999; Nurse and Philippson, 2003; see also Nurse, 1994–1995). Most of these classifications have in common that (1) they have some trouble defining an exact line between Narrow Bantu and closely related Bantoid languages in Cameroon, (2) they see a small set of languages (zones A, B, C, and bits of D and H, spoken in the northwest and north, in Cameroon, Gabon, Congo, and the northern fringe of the DRC, often called the Forest languages) as different from the rest, and (3) they divide the rest, the majority, into a smaller western (Angola, Namibia, parts of the DRC, Zambia, and Botswana) and a larger eastern set (all of eastern and most of southern Africa). Using nonlexical criteria, Nurse and Philippson (2003) differed somewhat in their view of classification (and history). While acknowledging the northwestern/northern grouping and the western group, they also saw a distinct northeastern group (Uganda, Kenya, and Uganda), but otherwise viewed the remaining languages as a group defined negatively by not sharing the innovations of the west, northwest, north, and northeast.

History

The entire Bantu area historically was covered in work by Vansina (1995), and Vansina (1990) and

Ehret (1998) dealt in detail with the northwest/north and east/south/southwest, respectively. An older, slightly outdated, view of the development of Bantu history is found in the work of Oliver (1966) and Phillipson (1977).

Though Ehret and Vansina disagreed on many details, the general outline is clear. Within Niger-Congo, Bantu is part of a grouping currently called Benue-Congo. The ancestors of the Benue-Congo, farmers, lived between what is now the Ivory Coast and western Cameroon, starting some seven millennia ago. By 3000 B.C., the ancestors of the Bantu had emerged and had already divided into what were later to become western and eastern Bantu. During the next millennium, they all moved slowly south and east across Cameroon, carrying the West African planting tradition with them. By 1000 B.C., they had moved much further into the rainforest and had reached various points on or near the Congo (Zaire) River in today's DRC, so that there was a wide range of Bantu communities in the forest (Vansina, 1990: 51–54). There is a popular myth that the huge equatorial rainforest is uninhabitable and uncrossable. In fact, today it has some 12 million inhabitants, spread across 450 ethnic groups (Vansina, 1990: 3), and the early Bantu crossed it easily, following the major rivers (the Congo (Zaire), Kasai, Sankuru, Lualaba, Lomami, and Sangha). By 1000 B.C., the ancestors of today's eastern Bantu were already at the eastern end of the forest, at the western edge of the Great Lakes region, to the east and south of the forest. This is necessarily a shortened and simplified version of events: in particular, it ignores the northwestern and northern Bantu communities, speaking the so-called Forest languages, whose history is somewhat separate and not further followed here.

Ehret saw the ancestors of today's eastern Bantu communities as having divided into two groups, an incipient northern and an incipient southern group, by 3000 years ago, both located in the area west of the West Rift valley in East Africa. The former group likely was to the west and south of Lake Victoria, the latter being west of Lake Tanganyika. The northern group had reached Lake Victoria by the middle of the first millennium B.C. During this period and later, these ancestors came across communities speaking Nilo-Saharan, Cushitic, and Khoisan languages, encounters that contributed to a diversified agriculture and boosted pastoralism. Iron working also had appeared in the region by this time, but its origins are disputed. During the next 500 years, some communities spread around Lake Victoria; some spread across Kenya and northern Tanzania to the coast by the early centuries A.D. and, by a couple of

centuries later, others spread south and southeast across Tanzania close to northern Mozambique.

Meanwhile, the southern offshoot of eastern Bantu had left the southern fringes of the rainforest and approached northeast Zambia by the second half of the last millennium B.C. They spread thence into much of Zambia, Malawi, Zimbabwe, southern Mozambique, and eastern South Africa, early Shona societies being established south of the Limpopo River by the third century A.D. These early groups ran across long-established Khoisan peoples in most of the region. Later movements, even into the second millennium A.D., resulted in the current configuration of Bantu communities in South Africa. Western Bantu likewise splintered. Sections moved east and northeast along the upper Congo River and its tributaries, and then southeast, so that nearly all of the rainforest was occupied by western Bantu populations by 1 A.D. Once the ancestors of a southern arm had crossed the lower Congo River and moved out of the rainforest into the adjacent savanna, during the latter half of the last millennium B.C., one section continued south across the Benguela Highlands in Angola and finally into northern Namibia, and another turned east and southeast and moved as far as western Zambia, along the Upper Zambezi River. Most western Bantu populations were in or near their current locations by the late centuries B.C. or the early centuries A.D.

With a few notable exceptions, most major movements of early Bantu-speaking peoples, east and west, were complete by the early centuries A.D., and the ancestors of most current Bantu populations had occupied central, eastern, and southern Africa by that time. Thereafter, some minor movements and local dispersals followed, with much contact, interaction, mixing, and assimilation.

Typology

This section sketches some characteristic features of Bantu phonology, morphology, and syntax, mainly those that occur widely but including a few less widespread, but intrinsically interesting. Exceptions to these generalities occur mainly in northwestern and northern languages.

Phonology

Nearly all Bantu languages have five or seven contrastive vowels, the few exceptions being mainly in Cameroon, Congo, and DRC, with languages with nine or more vowels. Five- and nine-vowel systems derive from earlier seven-vowel systems, the number usually assigned to Proto-Bantu. Despite the apparent similarity of the systems, phonetic realization varies

considerably, especially in languages with seven vowels. The full range of vowels is seen in most stem positions, whereas a reduced set occurs in other contexts, such as prefixes and extensions (the derivational suffixes following the stem). Vowel height harmony is widespread, whereby the historical degree-two vowels (the second highest vowels in the seven-vowel system) harmonize and lower to /e/ and /o/ after /e/ and /o/ in the root (and in some languages after /a/) (see Schadeberg, 1994/1995; Hyman, 1999, 2003; Kisseberth and Odden, 2003; Maddieson, 2003).

Proto-Bantu had a few pairs of words differentiated only by vowel length. Some languages have kept these, some have neutralized the length distinction, some have reintroduced it, and others have increased its function, often via loanwords. Phonetic vowel length is typically induced by vowel fusion, following prenasalized consonants, or gliding and 'compensatory lengthening' of the remaining vowel (in sequences of two vowels, /i, e/ and /u, o/ in the first position typically become the glides /y/ and /w/, respectively, before nonidentical vowels).

Most Bantuists credit Proto-Bantu with the following consonant system:

| | | | |
|---|---|---|---|
| p | t | c | k |
| b | d | j | g |
| m | n | ɲ | |

Here there is a voiced:voiceless contrast in plosives and a full set of nasals at nearly all the same points of articulation as the plosives. Most contemporary languages have these features, although the voiced plosives have often become continuants. Not shown here, there were also two sets of prenasalized consonants (mb, nd. . . : mp, nt. . .). Most languages still have the voiced set, but the voiceless set has been less stable over time, apparently because of the disparity in voicing between nasal and obstruent.

Some languages still have a simple consonant system that, although altered from the one shown here, derives from it fairly directly. Others have a much-expanded system, partly due to common phonetic processes such as palatalization, gliding, and voicing, but often resulting from a widespread process known as Bantu Spirantization. In this, the two high vowels in the original seven-vowel system affected the preceding plosives, typically producing affricates or fricatives: labial /pf, bv, f, v/ from /p, b/, labials from the nonlabial consonants before the high back vowel, and alveolar or palatal /ts, dz, s, z, etc./ from the nonlabial consonants before the front vowel. Typically, the two high vowels then merged with the degree-two vowels. The result was a smaller five-vowel inventory but a larger consonant inventory with voiced and voiceless plosives and fricatives.

Other consonant processes, geographically more limited, are defined by Dahl's Law and Meinhof's Law. Dahl's Law voices a voiceless stop if the obstruent in the next syllable is also voiceless (so Kikuyu *geki* from English 'cake'), which has interesting effects in long strings. Possibly linked to this is a more local phenomenon, Katupha's Law, which disallows aspirated consonants in adjacent syllables, deaspirating the first. Meinhof's Law affects sequences of noun, consonant, vowel, noun (consonant), or NCVN(C), deleting the first C, so *ngombe* 'cow' would become *gombe*. A local variant, the Kwanyama Law, produces the opposite result, *ngobe*. Syllables in Bantu are almost universally open, that is, CV or CVV. In restricted contexts, such as prefixes, other shapes (V, N, NCV) occur. A few languages, mostly in the northwest, may have closed syllables (CVC), due to loss of final vowels (but their tones are mostly kept).

Some 95% of Bantu languages are tonal and have a basic contrast between high (H) and low (L), or H and toneless. Contour tones (falling, rising) are usually restricted to bimoraic syllables. Downstepping of each successive H is common. Nouns and verbs show significant differences in the distribution of tone. Nominal prefixes are typically toneless and a number of stem patterns are possible, varying from language to language. Tones in verbs are more complicated than they are in nouns. Verb stems in many languages show a lexical contract between H and toneless, and affixes may also have their own tone, so that in some languages the tone of the verb is more or less the sum of individual tones, modified by certain general processes. In other languages, verb stems have no lexical tone, tone being assigned by general principles, often particular to certain tenses. Even languages with lexical stem tone often have grammatical tone, whereby an H may be assigned to a specific stem mora in certain tenses.

In many Bantu languages, the relationship between an underlying and a surface H is not direct, being modified by widespread principles and processes that favor or disfavor certain configurations. One such is tone spreading (tone of one syllable spreads to the next syllable(s), so being realized on two or more syllables), or tone shift (tone of one syllable is realized only on the next), typically from left to right. Another is avoiding situations whereby phonological structures – typically the intonational phrase or the word – end on an H. A third is the disfavoring of successive (nonsurface) H's, the obligatory contour principle (OCP). Working against the OCP is the plateau principle, whereby a toneless stretch between two H's is avoided. Finally, tones mark certain grammatical functions. Besides being associated with

certain tenses or groups of tenses, as already noted, tone often serves, for example, to distinguish statement from question, positive from negative, main from subordinate clause (the latter often including relative clauses), and the third person (H) from the second person singular subject prefix, which are otherwise segmentally identical.

Morphology

Bantu languages have the following word classes: noun, verb, pronoun, adjective (a small class; only a dozen or so are reconstructable for Proto-Bantu), numeral, demonstrative (often a three-way contrast), a small set of quantifiers and interrogatives, preposition (often a compound), and ideophone. One conjunction (*na*) is widespread (see Katamba, 2003; Nurse, 2003; Schadeberg, 2003).

Nouns consist of a stem and a prefix (L). Most prefixes have a CV-shape, stems are of the shapes -CV, -CVCV, -CVCVCV, etc., whereby the last vowel might be part of the stem or it might be a derivational suffix, so in Ha the final vowel of [umwáana] is part of the stem, whereas in *umuhanuuz-i* 'advisor,' *umukén-e* 'pauper,' and *igisus-o* 'example,' the final vowels are derivational suffixes (from *-hanuur-* 'tell', *-ken-* 'miss,' and *-sus-* 'resemble,' respectively). Nouns in many but not all languages have an augment (also called preprefix, initial vowel), consisting of a vowel that reflects the vowel of the prefix. It has various pragmatic and syntactic functions.

All nouns are assigned to a class. Classes have four characteristics: (1) each class has a nominal prefix, (2) there is extensive concord between the noun and the constituents of the noun phrase and the subject and object prefixes in the verb, (3) there are typical singular-plural class pairings, often called genders, and (4) and there is some semantic content to each class and gender. Concord, incidentally, is not always automatic – animacy, for example, can sometimes override automatic class agreement. Typical languages have between 15 and 21 classes and at least six genders, leaving some single classes with no plural pairing. The classes (Cl.) have been given conventional numbers (Cl. 1, 2, 3, 4, etc.) and some genders are widespread (e.g., 1/2, 3/4, 5/6, 7/8, 9/10, 11/10, 12/13, 14/6). Classes 15, 16, and 17 are locative classes and typically have a single member, or even no regular members, their prefixes being added to other nouns. Gender 12/13 (and sometimes 7/8) is a diminutive gender, gender 5/6 (and sometimes 20/21) is an augmentative. Gender 9/10 in many languages appears to act as a dumping ground for nouns that do not fit elsewhere. Nouns may be shifted from one (primary) class to another (derived), typically the

classes just mentioned, in which case the prefix and semantic features of the new class will be added to or replace those of the primary class. Thus there are the Ha words *u-mw-ána* 'child' (Cl. 1), *a-ka-ána* 'small child' (13), *u-tw-ána* 'small children' (13), *i-zí-iko* 'fireplace' (5), and *ku-zí-iko* 'to/on the fire' (16 + 5); also, in Haya (E22), there is *o-mu-ntu* 'person' (1), but *o-lu-ntu* 'tall, slim but slightly ridiculous person' (11). A few northwestern and northern languages have greatly reduced the number of noun classes to a handful, or even to none.

For many decades, it was maintained that with the exception of the derived genders and of gender 1/2 (humans), it was not possible to state the semantic content of most classes and genders, other than by listing typical and obvious groupings, and there were many anomalies. These groupings and anomalies occur across Bantu. Thus gender 3/4 typically contains plants, bushes, trees, and some natural phenomena, but it also widely contains 'year' and 'end,' and, in Swahili, 'mosque'! Contemporary attempts have been made to look at semantic content differently. Rather than trying to reduce content to one or a very few clearly storable characteristics, the new, cognitively inspired approach tries to find coherence in the notion of semantic networks, thus plants > objects made from plants > powerful things (e.g., medicine), or plants/trees > long, extended shape > time trajectory (e.g., 'year, journey'). This still leaves unexplained exceptions, but may lead to even better results when applied to more languages.

It is interesting to note in closing that the final semantic contrasts remaining in languages that have reduced their noun classes almost to zero are those of languages recognized as pidgins. Thus the Cameroonian language Kako (Katanga, 2003: 108), certain D30 languages in the northeastern DRC, and Pidgin Swahili, as spoken in Nairobi, have in common that they have only two or three classes left, retaining only the distinction animate/inanimate or human/nonhuman.

Bantu languages are verby, that is, the verb is not only the organizational center of the sentence but encodes more information than any other word class, information that in, for example, English requires several words. The verb structure is agglutinating and may include up to 20 morphemes in some languages (Nurse and Philippson, 2003c: 9). These two structures cover the main possibilities for the one-word verb:

NEG₁ - prefix - formative - object - root - extension - final vowel - postfinal
 prefix - NEG₂ - formative - object - root - extension - final vowel - postfinal

The only two obligatory constituents are root and final vowel, which cooccur in the imperative. Several morphemes may cooccur at prefix, formative, object, extension, and postfinal, typically in a canonical order. The structures differ only in the position of the NEG. Over the past two decades, phonologists have interpreted these linear structures as a hierarchy. Root and extension form the derivational stem: extensions are tonally neutral, have a canonical VC shape, and have a reduced five-vowel system; the derivational stem is the domain of vowel harmony with the root. Derivational stem and final vowel form the inflectional stem, the domain of reduplication, vowel coalescence, and limited consonant harmony with the final root consonant. Derivational stem and object form the macrostem, the domain of certain tonal phenomena. Finally, the macrostem combines with all preceding material to form the verbal word. This synchronic division of the verb into macrostem and prefixes corresponds well with likely historical development – other Niger-Congo languages have the macrostem, to which Bantu prefixes were added later.

Always or nearly always encoded in the inflected verb are subject, tense, aspect, mood, valency, and negation. Subject concord is usually obligatory and encoded at the prefix in both of the preceding structures, whether the subject noun is present or not. Tense is most often encoded at formative, less often at the final vowel or before the prefix. Bantu languages typically have multiple past and future reference: 83% of a database of 100 languages geographically representative of all 500 had between two and five discrete past tenses (40% had two, 32% had three), and 87% had one to three futures (46% had just one, 25% had two). Aspect seems to have been originally marked at the final vowel, but today also appears at formative: perfective, imperfective, progressive, habitual, anterior (also called ‘perfect’), and persistent are the commonest aspects. Mood is most often subjunctive, marked by a suffixal [e] at the final vowel. Valency changes are marked at extension and include causative, applicative (encompassing various functions), impositive, neuter/decausative, positional, reciprocal/plurational, repetitive, extensive, tentative, reverse, and passive. Negation appears variously; 51% of the database languages have two negatives, one associated with subordinate clauses, relative clauses, subjunctives, and imperatives, the other with main clauses. The former is typically but not always marked at NEG₂, the latter at NEG₁; 28% of the database languages have a single negative, either at NEG₁ or NEG₂ or pre- or postverbally, and 15% of the languages have more than two negatives. Tense, aspect, mood (TAM)

distinctions in negative verbs may differ from those in positives.

Less often, rarely, or not encoded in the verb are relative markers, focus, pronominal objects, and other categories. Relatives are most often marked before or at prefix in the second structure, and often the main marking is tonal. Focus can highlight several categories (e.g., the lexical verb itself, what follows the verb, or the aspect) and is usually indicated as a second or third morpheme in the formative slot, or verb initially. Pronominal object marking is also variable: some languages allow no object markers in the verb, some allow one, some allow two, and in a few languages four and even five have been recorded, especially in association with an applicativized verb.

A second or third morpheme in the formative slot marks consecutive, itive, or ventive in some languages. Many languages allow compound verbs, whereby the first verb is a tense-marked auxiliary, most often ‘be,’ and the second, lexical verb carries aspect. Many TAM markers are visibly grammaticalized, reduced forms of auxiliaries.

Syntax

Bantu languages belong to Heine’s (1976) Type A, having subject (S) (Aux)-verb (V)-object (O)-X, whereby there may be two objects (double object marking, rather than direct and indirect), and X represents adverbials (the Cameroonian language, Nén, with subject-object-verb (SOV), is the only known exception): prepositions: and noun phrase constituents, including relative clauses and the genitive construction, follow the head noun. The following Ha examples illustrate these and other features mentioned earlier:

- inkokó ziníni zóóse záanje
 ‘all my big chickens’ (lit. chickens big all my)
- izo inkokó ziníni zibíri
 ‘those two big chickens’
- igúnira dzuuzúye imbutó
 ‘bag which.is.full.of seeds’
- ubwáato bwa-dáatá
 ‘canoe of-father’
- ba-ø-teera ibiharagi
 ‘they-sow beans’ (postverbal focus)
- ba-ø-ra-téera
 ‘they sow’ (verbal focus)
- wari wágiye heéhe
 ‘where had you gone?’ (lit. you.were you.went where)
- kéera ha-rabáaye
 ‘once there-was’ (Class 16)

yasutse-mwó amáazi
 'she.poured-in water' (Class 16)

urondera-kó (Class 17)
 '...where you sought' (lit. you sought-where)

yamúhaaye umukáaté umwáana kumwóonga
 'she.gave bread to.child at.river'

The first four examples illustrate the order of constituents of the noun phrase. Harjula (2004: 131), from whom these examples come, stated that some of the constituents may change their order "without a change in the meaning" (in other languages, a change of place implies a change in emphasis) and that demonstratives precede the noun (in other languages, they may precede or follow). The fifth and sixth examples show one kind of focus contrast and one way of doing it: the form showing the close relationship between verb and postverbal constituent, also called the conjunctive, has a zero marker (ϕ) between subject marker and verb, whereas the form with focus on the verb, the disjunct, has a morpheme *ra* and retains the H of the stem. The seventh example shows a typical compound verb ('be' followed by main verb) and a wh-question: the wh-word typically retains the position of the element replaced, at least for nonsubjects. Yes/no questions are indicated either by a question marker at the beginning or end of the sentence, or by use of tone. Examples 8–10 show locatives in subject, object, and relative function, respectively, spatial relations being typically coded on the verb. The last example shows the ditransitive verb 'give' with two objects and an adverbial. Of this, Harjula said: "When there are two object prefixes the more indirect (i.e., the patient) is closer to the stem." This runs counter to Bearth's (2003: 127) claim that "the widespread tendency in Bantu languages is to assign the positions next to the verb on account of a hierarchy of parameters defined, in terms of (i) animacy of the referent (human > animate > inanimate), (ii) semantic role relationship (beneficiary > goal > patient > locative), (iii) participant category (first > second > third person), and (iv) number (plural > singular)". This is true of noun phrases following the verb, and their mirror image, object prefixes preceding it. Finally, although the canonical word order is SVO, considerable word-order variation is possible for pragmatic purposes. The position to the right of the verb, in particular, acts as a focus position.

Bibliography

Bastin Y, Coupeux A M & Mann M (1999). *Continuity and divergence in the Bantu languages: perspectives from a lexicostatistic study*. Tervuren: Musée de l'Afrique Centrale.

- Bearth T (2003). 'Syntax.' In Nurse & Philippson (eds.). 121–142.
- Bernsten J (1998). 'Runyakitara: Uganda's 'new language.' *Journal of Multilingual and Multicultural Development* 19(2), 93–107.
- Ehret C E (1998). *An African classical age*. Charlottesville: University Press of Virginia and London: James Currey.
- Ehret C E (1999). 'Subclassifying Bantu: stem-morpheme innovations.' In Hombert & Hyman (eds.). 43–148.
- Guthrie M (1948). *The classification of the Bantu languages*. London: International African Institute.
- Guthrie M (1971). *Comparative Bantu* (vol. 2). Farnborough: Gregg International Publishers.
- Harjula L (2004). *The Ha language of Tanzania*. Cologne: Rüdiger Köppe Verlag.
- Heine B (1976). *A typology of African languages based on the order of meaningful elements*. Berlin: Dietrich Reimer.
- Heine B & Nurse D (eds.) (2000). *African languages*. Cambridge: Cambridge University Press.
- Hombert J-M & Hyman L M (eds.) (1999). *Bantu historical linguistics: theoretical and empirical perspectives*. Stanford: Center for the Study of Language and Information.
- Hyman L H (1999). 'The historical interpretation of vowel harmony in Bantu.' In Hombert & Hyman (eds.). 235–295.
- Hyman L M (2003). 'Segmental phonology.' In Nurse & Philippson (eds.). 42–58.
- Katamba F (2003). 'Bantu nominal morphology.' In Nurse & Philippson (eds.). 103–120.
- Kisseberth C W & Odden D (2003). 'Tone.' In Nurse & Philippson (eds.). 59–70.
- Maddieson I (2003). 'The sounds of the Bantu languages.' In Nurse & Philippson (eds.). 15–41.
- Maho J (2000). 'A classification of the Bantu languages: an update of Guthrie's referential system.' In Nurse & Philippson (eds.). 639–651.
- Nurse D (1994–1995). '"Historical" classifications of the Bantu languages.' *Azania* 29, 65–89.
- Nurse D (1999). 'Towards a classification of East African Bantu languages.' In Hombert & Hyman (eds.). 1–41.
- Nurse D & Philippson G (eds.) (2003a). *The Bantu languages*. London and New York: Routledge.
- Nurse D & Philippson G (2003b). Nurse & Philippson (eds.) 'Towards a historical classification of the Bantu Languages.' 164–181.
- Nurse D & Philippson G (2003c). 'Introduction.' In Nurse & Philippson (eds.). 1–12.
- Oliver R (1966). 'The problem of the Bantu expansion.' *Journal of African History* 7, 361–376.
- Philippson D W (1977). *The later prehistory of eastern and southern Africa*. London: Heinemann.
- Schadeberg T C (1994/1995). 'Spirantization and the 7-to-5 vowel merger in Bantu.' *Belgian Journal of Linguistics* 9, 73–84.
- Schadeberg T C (2003). 'Derivation.' In Nurse & Philippson (eds.). 71–89.

Sommer G (1992). 'A summary of language death in Africa.' In Brenzinger M (ed.) *Language death: factual and theoretical explorations with special reference to East Africa*. Berlin: Mouton de Gruyter. 310–418.

Vansina J (1990). *Paths in the rainforest*. Madison: The University of Wisconsin Press.

Vansina J (1995). 'New linguistic evidence and the 'Bantu expansion.' *Journal of African History* 36, 173–195.

Wolff H E (2000). 'Language and society.' In Heine & Nurse (eds.), 298–347.

Bashkir

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Bashkir (*bašqort tēlē*, *bašqořta*) belongs to the northern group of the northwestern, or Kipchak, branch of Turkic. Its main area of distribution is the basin of the Belaya River and the southwestern slopes of the Ural Mountains. The Republic of Bashkortostan, or Bashkiria (*Bašqortostan Respublikahı*), which belongs to the Russian Federation and whose capital is Ufa (Öfö), borders on Tatarstan, the Udmurt Republic, and the Orenburg, Perm, Sverdlovsk, and Chelyabinsk regions. Of the more than 4 million inhabitants of the Republic, Bashkirs make up only 22%. Other groups include Russians, Tatars, Chuvash, Udmurts, Mari, and Ukrainians. Bashkir-speaking groups are also found south of Kuybyshev and east of Ural, in the regions Orenburg, Chelyabinsk, Samara, Kurgan, and Sverdlovsk. The total number of speakers of Bashkir is about 1.4 million.

Origin and History

The Bashkirs previously lived farther to the east, in West Siberia, first as subjects of the Volga Bulgar state and, after 1236, under Mongol rule. They reached their present-day territory under the Golden Horde. With the disintegration of the Golden Horde, the Bashkir territory was divided between the three khanates of Kazan, Noghay, and West Siberia. Bashkirs and Tatars came under Russian rule at the end of the 18th century. In 1919, a Bashkir Autonomous Soviet Socialist Republic was established. In 1992, Bashkortostan became an autonomous republic within the Russian Federation.

Related Languages and Language Contacts

Bashkir is closely related to Tatar and constitutes a connecting link to Kazakh. The different origins of

its speakers are reflected in heterogeneous linguistic features. Since Bashkir and Tatar varieties have been in close contact for many centuries, the boundaries between them are not always clear.

The Written Language

The Bashkirs used a local variety of Chagatay as their written language until the beginning of the 20th century, when they adopted written Tatar. A Bashkir standard language, mainly based on the eastern (Kuvakan) dialect, was established in the Soviet era. The Arabic script was replaced in 1929 and 1930 by a Roman-based script. The Cyrillic-based script system that was introduced in 1939 and 1940 differs considerably from the script of the Tatar system.

Distinctive Features

Bashkir exhibits most linguistic features typical of the Turkic family (see **Turkic Languages**). It is an agglutinative language with suffixing morphology and a head-final constituent order (subject-object-verb). In the following discussions, some of the distinctive features of Bashkir will be dealt with, with focus in particular on certain comparisons with Tatar.

Phonology

The Bashkir vowel system is very similar to that of the Tatar system. It comprises fully articulated and reduced vowels and exhibits the same systematic vowel shifts. Thus, low vowels of the first syllable have been raised: *e* > *i* (*bin* 'you' (<*sen*)), *o* > *u* (*yul* 'way' (<*yol*)), *ö* > *ü* (*hüd* 'word' (<*söz*)). High vowels have been centralized and reduced: *i* > *ě* (*těđ* 'knee' (<*tiz*)), *u* > *ö* (*moron* 'nose' (<*burun*)), *ü* > *ö* (*kön* 'day' (<*kün*)).

In its consonant system, Bashkir differs from Tatar and approaches Kazakh. Thus, *č* has developed to *s* (*kis* 'evening' (Tatar *kič*), *sės* 'hair' (Tatar *čěč*)). Word- and suffix-initial *s* has developed to *h* (*hari* 'yellow' (Tatar *sari*), *bul-ha* 'if it is' (Tatar *bul-sa*)). In other cases, *s* has developed to *θ* (*ki θ* 'to cut'

(Tatar *kis*). The corresponding voiced sibilant *z* has developed into *δ* (*δur* ‘big, great’ (Tatar *zur*), *hüδ* ‘word’ (Tatar *süz*)). Interdental sibilants are also typical of Turkmen. Bashkir exhibits word-initial *y-* in cases in which Tatar has *ǰ* (*yǰlı* ‘warm’ (Tatar *ǰılı*)). This phenomenon also affects old loanwords, e.g., *yen* ‘soul’ (< Persian *ǰān*, cf. Tatar *ǰan*). In its vowel harmony system, Bashkir is similar to Turkmen, Kirghiz, and some other languages in that low suffix vowels are rounded after a rounded vowel in the preceding syllable, e.g., *bölöt* ‘cloud’ (Tatar *bölit*), *öñ* ‘for’ (Tatar *öñen*).

The rules for consonant assimilations are much more complicated than they are in Tatar. Suffix-initial consonants may have up to four variants (plural *qala-lar* [city-PL] ‘cities,’ *at-tar* [horse-PL] ‘horses,’ *kül-der* [lake-PL] ‘lakes,’ and *taw-δar* [mountain-PL] ‘mountains,’ or ablative *qala-nan* [city-ABL] ‘from the city,’ *taw-δan* [mountain-ABL] ‘from the mountain,’ *at-tan* [horse-ABL] ‘from the horse,’ and *yalan-dan* [steppe-ABL] ‘from the steppe’). The third-person personal pronouns are singular *ul* ‘he, she, it’ and plural *ular* ‘they’ (Tatar *ul*, *alar*). The oblique stem of *ul* is *un-* (Tatar *an-*). The demonstrative pronouns include *bıl*, *bıñaw* ‘this,’ *oşo* ‘this here,’ and *şul*, *ul*, *anaw*, *tęę* ‘that.’

Dialects

Bashkir has a few main dialects and numerous sub-dialects. The eastern or mountain (Kuvakan) dialect comprises the subdialects Ay, Argayash, Salyut,

Miyas, and Kizil. The southern (Yurmat) group comprises Ik-Sakmar and the central dialect group comprises Kara-Idil and Dim. There are important differences between the eastern and southern dialects. The steppe, or southwestern, dialects have been strongly influenced by Tatar.

Bibliography

- Benzing J (1959). ‘Das Baschkirische.’ In Deny J *et al.* (eds.) *Philologiae turcicae fundamenta* 1. Aquis Mattiacis: Steiner. 421–434.
- Berta Á (1998). ‘Tatar and Bashkir.’ In Johanson L & Csató É Á (eds.) *The Turkic languages*. London & New York: Routledge. 283–300.
- Dmitriev N K (1927). ‘Étude sur la phonétique Baschkire.’ *Journal Asiatique* 210, 193–252.
- Juldaşev A A (ed.) (1981). *Grammatika sovremennogo baškirkogo literaturnogo jazyka*. Moskva: Nauka.
- Maksjutova H X (1976). *Vostočnyj dialekt baškirkogo jazyka v sravnitel’no-istoričeskom osveščanii*. Moskva: Nauka.
- Miržanova S F (1979). *Južnyj dialekt baškirkogo jazyka*. Moskva: Nauka.
- Poppe N (1964). *Bashkir manual. Descriptive grammar and texts with a Bashkir–English glossary* (Indiana University Publications, Uralic and Altaic Series 36). Bloomington/The Hague: Mouton & Co.

Relevant Website

<http://www.turkiclanguages.com> – Website with many Turkish language resources.

Basque

J I Hualde, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Basque (Euskara) is the only remaining vestige of the linguistic situation in western and central Europe before the Indo-European expansion. Although many attempts have been made to relate Basque to other languages of the world, none of them is generally considered to have been successful. Genetic links with the Finno-Ugric family, the languages of the Caucasus, or any other living language for which some scholars have sought a genetic relationship with Basque would be so remote that no solid proof is likely to emerge.

As for the extinct language of the ancient Iberians, once spoken along the Mediterranean coast of Spain and known to us from a relatively large number of inscriptions, the fact that Basque has been of little help in deciphering these inscriptions forces us to discard the hypothesis that the two languages are closely related (although they do share a number of phonological and morphological features, attributable to areal phenomena). Basque is thus a language isolate.

Throughout its known history, Basque has been spoken in an area of variable extent on both sides of the western Pyrenees and along the coast of the Bay of Biscay. The present-day Basque-speaking area (*Euskal Herria*) corresponds to parts of three different administrative units, two in Spain and one in

France. There are currently approximately 700 000 speakers of Basque, almost all fully bilingual in either Spanish or French. The largest number of speakers is found in the Autonomous Community of the Basque Country (ACBC), which comprises the provinces of Bizkaia, Gipuzkoa, and Araba (in Basque)/Alava (in Spanish). Here Basque is co-official with Spanish and has an important presence in the educational system. In this region, the number of Basque speakers is growing in areas where the language is natively spoken by part of the population as well as in other areas, such as the city of Bilbao and most of Araba/Alava, where the Basque language had been lost centuries ago. Basque also enjoys some official recognition in Navarra (in Basque, Nafarroa), which is a separate autonomous community within the administrative structure of Spain. Although the greatest part of Navarra was Basque-speaking just a few centuries ago, the language suffered a strong geographical recession in the 19th and 20th centuries, and nowadays it is spoken natively only in the northwestern area of this region. The French Basque country comprises approximately the western half of the Département des Pyrénées Atlantiques. In most of the Basque-speaking area of France, the transmission of the language has seriously declined in the last few decades. Only a small percentage of children are currently learning Basque in this area.

From toponyms and other sources, we know that historically the Basque language was spoken over a larger area. The word *Basque* derives from Vascones, a nation that in Roman times occupied most of Navarra and northern Aragon. Across the Pyrenees, there is abundant epigraphic evidence showing that Basque or a very similar language was also spoken in the territory of the Aquitani. The Aquitanian inscriptions on tombstones, in Latin, provide only evidence for proper names, but they contain such clear elements as *ANDERE* (cf., Basque *andere* 'woman') and *CISSON* (cf., Basque *gizon* 'man') as proper names for individuals of the respective sex. For the late Middle Ages, we have documentary evidence that Basque was spoken both in northern Aragon and in areas of La Rioja and northern Castile, to the west and south of the present-day Basque country. It is, however, likely that the historical presence of the Basque language in these latter areas, and perhaps even in part of the territory of the ACBC, is due to territorial expansion during the early Middle Ages.

One reason for the hypothesis that Basque may have occupied a compact area at some point after the fall of the Roman Empire is that dialectal diversity within Basque is relatively small and clearly not ancient. Many obvious innovations are shared by all dialects. In some aspects, such as the accentual system

and the morphology of finite verb forms, variation is, nevertheless, considerable (even if it is due to relatively recent diversification) and in fact virtually every valley or town has a recognizable local variety. *Euskara batua* (unified Basque), the standard promoted by the Basque Academy, which is based on the literary tradition of central areas both to the north and the south of the Pyrenees, has been enormously successful in its social implantation through its use in the educational system and in the media.

Most Basque dialects have five vowel phonemes /i e a o u/. Zuberoan (Souletin) and a few other varieties spoken in France have a sixth oral vowel /y/ as well as contrastively nasalized vowels. A common consonantal inventory, such as is found in Gipuzkoan, is the following (the most common orthographic representation follows in parentheses when it is different from the phonetic symbol): /p t c (tt) k b d ʒ (dd) g t͡s (tz) t͡ʃ (tx) ʃ (z) ʂ (s) ʎ (x) x (j) m n ɲ (ñ) l ʎ (ll) r (r) r (rr)/. The most unusual aspect of this inventory is presented by the contrast between the two fricatives, lamino-alveolar *z* /ʃ/ (*izan* 'be') and apico-alveolar *s* /s/ (*esan* 'say'), and the two corresponding affricated segments (*atzo* 'yesterday,' *atso* 'old woman'). All Bizkaian and some Gipuzkoan varieties have lost this contrast. The phoneme /x/ is found only in the speech of speakers from Spain. Besides being found in borrowings from Spanish, in central areas (Gipuzkoa and some neighboring regions) it also appears in native words as a result of an evolution /j > ʒ > ʃ > x/ (like in Castilian Spanish). Other dialects have stopped at various stages along this evolutionary path. The result is that orthographic *j* in native words such as *jan* 'eat' is subject to much variation in its pronunciation. Whereas the official standard pronunciation is /jan/, the Gipuzkoan form /xan/ is also in widespread usage in standard Basque, and locally forms like /ʒan/ and /jan/ are also used. Conversely, a phoneme /h/ (orthographic *h*: *hemen* 'here,' *aho* 'mouth') is used only in parts of the French Basque country. That is, for most speakers orthographic *h* is silent. The (pre)palatal consonants have a special status. One way to form diminutive/affective forms is by palatalization, for example, *tanta* 'drop,' *ttantta* /canca/ 'small drop,' *zezen* 'bull,' *xexen* /ʃeʃen/ 'little bull.' A pitch-accent system strikingly similar to that of Tokyo Japanese, with a lexical contrast between accented and unaccented words, is found in the northern Bizkaian area. The most common accentual system (in Gipuzkoan and neighboring areas), however, has regular stress on the second syllable.

Marking of grammatical functions works on a strictly ergative basis, with one case (absolutive, morphologically unmarked) assigned to objects and intransitive subjects and another (ergative, *-k*) assigned

to transitive subjects: *lagunak liburua dakar* 'the friend is bringing the book,' *laguna dator* 'the friend is coming.' Nevertheless, a class of syntactically intransitive verbs takes ergative subjects (and transitive auxiliaries and agreement) in a somewhat unpredictable manner: *lagunak dantzatu du* 'the friend has danced.' Finite verb forms are marked for agreement with up to three arguments (subject, object, and indirect object): *dakarzkiguzu* 'you (-zu) are bringing them (-z-) to us (-gu)' (-kar- is the verb root 'bring'; -ki- is a dative pre-prefix). In addition, in the familiar treatment, an addressee who is not an argument of the verb is also obligatorily encoded in the morphology of verbs in main clauses. Thus, for instance, plain/formal *dakit* 'I know it' is replaced, in the familiar treatment, by *zekiat* 'I know it (male addressee)' or *zekinat* 'I know it (female addressee).'

Although both SOV and SVO orders are common in texts, verb-final structures are more basic: *gizona da* 'it is the man.' Focalized elements and question words are normally immediately preverbal. Main verbs precede auxiliaries (*etorri da* '(she/he) has come'), except in negative clauses (*ez da etorri* '(she/he) has not come').

Articles and demonstratives are phrase-final: *laguna* 'the friend,' *lagun bat* 'a/one friend,' *lagun hori* 'that friend,' *lagun gazte hori* 'that young friend.' Although, as shown in the last example, adjectives follow nouns, genitives and relative clauses precede the head noun, as in most other SOV languages (*lagunaren liburua* 'the friend's book,' *etorri den laguna* 'the friend who has come'). Noun phrases are inflected for number and case by suffixes attached to the last word in the phrase: *lagunari* 'to the friend,' *lagun onari* 'to the good friend,' *etorri den lagun gaztearentzat* 'for the young friend who has come.'

Bibliography

- Allières J (1972). *Manuel pratique de basque*. Paris: A. & J. Picard.
- Aulestia G & White L (1992). *Basque-English English-Basque dictionary*. Reno: University of Nevada Press.
- Euskaltzaindia [Academy of the Basque Language] (1985). *Euskal gramatika: Lehen urratsak* [Basque grammar: first steps] (5 vols). Bilbao, Spain: Euskaltzaindia.
- Goenaga P (1980). *Gramatika bideetan* [On the paths of grammar]. Donostia, Spain: Erein.
- Gorochategui J (1984). *Onomástica indígena de Aquitania*. Bilbao, Spain: University del País Vasco.
- Hualde J I (1991). *Basque phonology*. London: Routledge.
- Hualde J I & Ortiz de Urbina J (eds.) (1993). *Generative studies in Basque linguistics*. Amsterdam: John Benjamins.
- Hualde J I & Ortiz de Urbina J (eds.) (2003). *A grammar of Basque*. Berlin: Mouton de Gruyter.
- Hualde J I, Lakarra J A & Trask R L (1995). *Towards a history of the Basque language*. Amsterdam: John Benjamins.
- King A R (1994). *The Basque language: a practical introduction*. Reno: University of Nevada Press.
- King A R & Olaizola Elordi B (1996). *Colloquial Basque: a complete language course*. London: Routledge.
- Lafitte P (1979). *Grammaire basque (navarro-labourdin littéraire)* (3rd edn.). Donostia, Spain: Elkar.
- Michelena L (1985a). *Fonética histórica vasca* (3rd edn.). Donostia-San Sebastián, Spain: Diputación Foral de Gipuzkoa.
- Michelena L (1985b). *Lengua e historia*. Madrid: Paraninfo.
- Michelena L (1987). *Palabras y textos*. Vitoria-Gasteiz, Spain: Universidad del País Vasco.
- Michelena L (1988a). *Historia de la literatura vasca* (2nd edn.). Donostia-San Sebastián, Spain: Erein.
- Michelena L (1988b). *Sobre historia de la lengua vasca* (2 vols). Supplements of Anuario del Seminario de Filología Vasca 'Julio de Urquijo' 10. Donostia-San Sebastián, Spain: Diputación Foral de Gipuzkoa.
- Ortiz de Urbina J (1989). *Parameters in the grammar of Basque*. Dordrecht: Foris.
- Patrick J D & Zubiri I (2001). *A student grammar of Euskara*. Munich: Lincom-Europa.
- Rebuschi G (1977). *Essais de linguistique basque*. Supplements of Anuario del Seminario de Filología Vasca 'Julio de Urquijo' 35. Donostia-San Sebastián, Spain: Diputación Foral de Gipuzkoa/Bilbao, Spain: Universidad del País Vasco.
- Rijk R P G de (1998). *De lingua vasconum: selected writings*. Supplements of Anuario del Seminario de Filología Vasca 'Julio de Urquijo' 43. Bilbao, Spain: Universidad del País Vasco/Donostia-San Sebastián: Diputación de Gipuzkoa.
- Sarasola I (1982). *Historia social de la literatura vasca*. Madrid: Akal.
- Trask R L (1977). *The history of Basque*. London: Routledge.
- Villasante L *Historia de la literatura vasca*. Aránzazu, Spain: Editorial Aránzazu.

Belorussian

S Young, University of Maryland Baltimore County,
Baltimore, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Belorussian (*belaruskaja mova*; Belarusian, Belarusan), which together with Ukrainian and Russian forms the East Slavic branch of the Slavic languages, is the native language of some 8 million speakers in the Republic of Belarus. The standard language is based on the central dialect of the Minsk region. In an earlier form known as Old Belorussian, West Russian, or among contemporaries simply as *rus'skij*, Belorussian served from the 15th through the late 17th centuries (when it finally yielded to Polish) as the chancery language of the multiethnic Grand Duchy of Lithuania (which in 1569 became part of the Polish Commonwealth). Thereafter, with political bans on publication in the language, Belorussian went into a period of decline. It was not until the first decades of the 20th century that Belorussian experienced a revival, with roots not in the distant literary traditions of the Grand Duchy, but in the vernacular of the countryside. The first legal Belorussian periodical, *Naša Niva* 'Our Cornfield' (1906–1915), attracted contributions from leading intellectuals of the day and did much to promote structural and orthographic uniformity in the language. The first attempt at a normative grammar of the language was Branislaŭ Taraškevič's *Belaruskaja hramatyka dlja škol* 'Belorussian grammar for schools' (1918). The consolidation of grammatical norms continued well into the 20th century.

Belorussian, which is written in the Cyrillic alphabet, shares a number of phonological features with both Russian and Ukrainian. As in standard Russian, unstressed *o* is pronounced *a* (*ákanne*), and (as in certain Russian dialects) unstressed *e* becomes *'a* (*jákanne*). Unlike Russian, these features are reflected in the orthography (in the case of *jákanne*, only in pretonic position), which is set up on the phonemic, rather than morphophonemic, principle: *nažý* 'knives' (sg. *nož*) and *zjamljá* 'world' (pl. *zémli*). Most consonants occur in phonemically opposed palatalized–nonpalatalized pairs. East Slavic *tʲ* and *dʲ* have

assibilated to *tsʲ* and *dzʲ*: *dzéci* ['dzʲetsʲi] 'children' (Rus. *děti* ['dʲetʲi]); palatalized *rʲ* has been lost: *rad* 'row' (Rus. *rjad*). As in Ukrainian, the palatal affricates *č* and *šč* are pronounced hard, East Slavic *g* is a fricative [ɣ], and *v* becomes [w] (in transcription from Cyrillic, *ŭ*) in closed syllables: *halóŭka* 'head, dim.' (*halavá* 'head').

Morphological characteristics of the noun include the loss of a distinct neuter plural: *aknó* 'window' (pl. *vókny*; Rus. *oknó*, *ókna*); the alternation of stem-final velars and dental affricates in certain case forms: nom. sg. *ruká* 'hand' (dat. sg. *rucé*); and a tendency toward the spread of the first declension genitive plural marker *-oŭ* (unstressed *-aŭ*) to other declensions: *zímaŭ* (Rus. *zim*) 'of winters'.

The verb has two regular conjugation patterns, illustrated in the present tense by *něsci* 'to carry' (I) and *rabíc* 'to do, make' (II): 1SG *njasú*, *rabljú*; 2SG *njasěš*, *róbiš*; 3SG *njasé*, *róbic*; 1PL *nesëm*, *róbim*; 2PL *nesjacé*, *róbice*; 3PL *njasúc*, *róbjac*. Like Ukrainian, but unlike Russian, the third-person ending (lacking in the singular of pattern I) is palatalized. As in Ukrainian, there is a change of the masculine past tense marker *l* to *w*: *znaŭ* masc. 'knew' (fem. *znála*).

To a greater extent than in Ukrainian, the lexicon reflects the historical influence of Polish, chiefly from the period of the Polish-Lithuanian Commonwealth. Since the late 18th century unification with Russia, the influence of Russian has prevailed.

Bibliography

- Mayo P (1993). 'Belorussian.' In Comrie B & Corbett G (eds.) *The Slavonic languages*. London & New York: Routledge. 887–946.
- McMillin A (1980). 'Belorussian.' In Schenker A & Stankiewicz E (eds.) *The Slavic literary languages, formation and development*. New Haven: Yale Concilium on International and Area Studies. 105–117.
- Sherech Y [Shevelov G] (1953). *Problems in the formation of Belorussian*. New York: Linguistics Circle of New York.
- Stang C (1935). *Die Westrussische Kanzleisprache des Grossfürstentums Litauen*. Oslo: Jacob Dybwad.
- Wexler P (1977). *A historical phonology of the Belorussian language*. Heidelberg: C. Winter.

Bengali

H-R Thompson, School of Oriental and African Studies, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Bengali is the official language of Bangladesh and of the state of West Bengal in India. There is some controversy about the correct name of the language. The term 'Bangla' is increasingly in use, particularly among Indian linguists, for whom the term 'Bengali' may be associated with British India. It is likely that in the not-too-distant future 'Bangla' will replace 'Bengali.' With a total number of about 260 million speakers, Bengali is the world's fifth largest language.

Bengali, together with Assamese and Oriya, belongs to the eastern branch of Indo-Aryan languages. A high percentage of vocabulary is derived from Sanskrit, with lesser influences from Persian, Arabic, and English. Bengali has a very large vocabulary but the language situation is diglossic. The vocabulary used in spoken language is distinct from the highly Sanskritized words used in some literature and formal contexts. Many words have both a Sanskritic and a colloquial version, e.g., হস্ত /hōsto/, হাত /hat/ 'hand,' চন্দ্র /cōndro/, চাঁদ /cād/ 'moon,' দন্ত /dōnto/, and দাঁত /dāt/ 'tooth.' The early 20th-century rivalry between the *sadhu bhasha* (literary language) and *calit bhasha* (colloquial language) is now a thing of the past. Standard Colloquial Bengali, based on the language spoken in Kolkata, is the accepted norm. Some Bengali dialects retain the *sadhu* extended verb forms, e.g., আমি যাইতেছি /ami yaitechhi/ rather than the contracted *calit* form (আমি যাচ্ছি /ami yacchi/ for 'I am going.'

Dialects vary in phonological and grammatical deviation. Sylheti, the dialect spoken by most Bangladeshis living in the United Kingdom, has a high percentage of Persian words and is considered by some to be a separate language.

Orthography and Phonology

Bengali is written in a variant of the Devanagari script, which is related to but distinct from the script used for Sanskrit and Hindi. Writing is from left to right and is syllabic. There are 12 vowels or diphthongs, two semivowels, and almost 40 consonants. Bengali has a great number of conjunct letters that combine, in one symbol, two or more consonants or consonant-vowel clusters. Vowel signs are attached to consonants except at the beginning of words and syllables, where the full vowel is written.

An inherent vowel (pronounced /ɔ/ or /o/) is often pronounced when no other vowel is given.

Bengali, like other South Asian languages, distinguishes between aspirated/unaspirated and dental/palatal sounds. Nasalization occurs in individual words and is phonemic (কাদা /kada/ 'mud,' কাঁদা /kāda/ 'weep,' বাধা /badha/ 'obstruction,' বাঁধা /bādha/ 'bind') and in the distinction between ordinary and honorific personal pronouns, as in ওর /or/ 'his/her' (familiar) and ওঁর /ōr/ 'his/her' (honorific). Bengali spelling retains some Sanskrit features, but its pronunciation has evolved and changed. The word for 'soul,' though it is spelled আত্মা /atma/, is pronounced /atta/. The Sanskrit word for heaven 'swarga' becomes স্বর্গ, pronounced /sɔrgo/. The distinction between long and short u and i, which is present in the script, is no longer felt in pronunciation. Long /o/ can be represented by the vowel sign or by the inherent vowel. There are three symbols for the sound /ng/: ঙ, ঙ্গ, and the conjunct ঙ্গ. Their uses are to some extent interchangeable, but ঙ is never followed by a vowel, thus we have বাংলা /bangla/ (the name for the language), but বাংলালি /bangali/ (the adjective and name for the people). There are three sibilants in Bengali স /s/, শ /ʃ/, and ষ /ʂ/. Their pronunciation is /sh/, except in some conjuncts, in which it changes to /s/, e.g., বিশ্রাম /biʃram/ 'rest,' স্থান /sthan/ 'place,' and নাস্তা /nasta/ 'breakfast.'

Morphology and Syntax

Basic word order is subject-object-verb, but sentence parts can move freely to express emphasis. Bengali has a complex relative-correlative system – i.e., subordinating conjunctions such as যখন /yokhon/ 'when' and যদি /yodi/ 'if' almost invariably have a correlative conjunction in the main clause. Subordinate clauses generally precede main clauses.

Nouns have no grammatical gender. There are four cases, nominative, genitive, object, and locative. The **nominative** is unmarked. Number and definiteness is marked by determiners that are suffixed to nouns, but their use is partly defined by the context. Plural markers for animate and inanimate nouns are distinct from one another. All case endings are added after these suffixes, e.g., মেয়ে /meɣe/ 'girl,' মেয়েটি /meɣe-ti/ 'the girl,' and মেয়েটিকে /meɣe-ti-ke/ 'to the girl.'

In the **genitive** nouns, add র /r/ or এর /er/: বাবা /baba/ 'father,' বাবার /baba-r/ 'father's,' উকিল /ukil/ 'lawyer,' and উকিলের /ukil-er/ 'the lawyer's.' The genitive has a wide variety of uses, including possession (রিমার ভাই /rima-r bhai/ 'Rima's brother'), attribute (প্রেমের গল্প /præm-er golpo/ 'love story'), function (বসার ঘর /bosar ghor/ 'sitting room'), measurement (দুই ঘণ্টার ছবি /dui

ghoṇṭa-r chobi/ 'a film lasting two hours'), and cause or origin (সমস্যার সমাধান /somoṣṣa-r sōmadhan/ 'solution to the problem'). The genitive usually functions as the logical subject in impersonal structures.

The **object case** is marked by কে /ke/: from (বাবা /baba/ 'father' বাবা-কে /baba-ke/ 'to father'). The case ending is used to mark direct or indirect objects. The case marking is usually omitted for inanimate nouns, but can be added for emphasis or to avoid ambiguity.

The **locative** ending is এ /e/ after consonants: সহর /sōhor/ 'town' সহরে /sōhor-e/ 'in the town'; য় /y/ after আ /a/ (ঢাকা /dhaka/ ঢাকায় /dhaka-y/ 'in dhaka,' and তে /te/ after all other vowels: বালু /balu/ বালুতে /balu-te/ 'in the sand.' The locative is used to indicate place: বাড়িতে /baṛi-te/ 'at home,' direction: ঘরে /ghor-e/ 'into the house,' time: দশটায় /dɔʃta-y/ 'at ten o'clock,' cause: তার বলায় /tar bōla-y/ 'because of what he said,' instrument: হাতুড়িতে /hatuṛi-te/ 'with a hammer,' or origin: চেষ্টায় /ceṣṭa-y/ 'from/through trying.' The locative is rarely used with animate nouns.

Bengali has personal, demonstrative, relative, interrogative, and indefinite **pronouns**. Personal pronouns distinguish three grades of familiarity in the second person and two grades of respect in the third person. They distinguish singular and plural, but not gender. There is a three-way deictic distinction (*here*, *there*, and *removed from context*) that applies to third-person pronouns, attributive adjectives, demonstratives, and place adverbials, for instance, এ মেয়ে /e meye/ 'this girl,' ও মেয়ে /o meye/ 'that girl (over there),' সে মেয়ে /se meye/ 'that girl, (removed from context),' এখানে /ekhane/ 'here,' ওখানে /okhane/ 'there,' and সেখানে /sekhane/ 'in that place.'

Adjectives precede nouns and are indeclinable. For comparisons, auxiliary words are used:

- (1) আমার ভাই আমার চেয়ে লম্বা।
/amar bhai amar ceṣe lōmba/
my brother my than long
'My brother is taller than me.'
- (2) এই গাছ সবচেয়ে সুন্দর।
/ei gach sobceṣe sundor/
this tree all than beautiful
'This is the most beautiful tree.'

Postpositions are, with a few exceptions, noun forms: about my parents: *on the subject of my parents*; or verbal participles: with the hammer: *having taken the hammer*.

Verb conjugation is very regular. Verb endings are the same for singular and plural. Some active verbs can be extended to form causative verbs, e.g., জানা /jana/ 'know' becomes জানানো /janano/ 'inform'; দেখা /dækha/ 'see' becomes দেখানো /dækhanō/ 'show'. There are, morphologically, eight tenses. Present and past tense have simple and progressive

aspect. Perfect tenses (present and past) can express not only perfective aspect but are also used to refer to past events or actions directly. The past habitual is used for remote past events and for subjunctive uses. The future tense forms the after-state of all other tenses. Tense use is much freer than in English – in fact, narrative texts gain color and liveliness through frequent tense changes.

Every verb has four **nonfinite verb forms**: infinitive, verbal noun, conditional, and perfective participle. Conditional and perfective participles, in particular, offer in very concise forms a great range of meanings. The conditional participle is formed by adding লে /le/ to the stem of the verb: থাকলে /thakle/ from থাক- /thak-/ 'stay.' It can be used temporally as well as conditionally and its temporal structure is determined by the main clause, thus a phrase such as সে থাকলে /se thakle/ has a range of meaning, from 'when he is here' to 'if he were alive.' The perfective participle, formed by adding এ /e/ to the verb stem, describes in its basic use a preceding action (e.g., খবরটা শুনে সে বাইরে গেল। /khōborṭa ṣune se baire gælo/ 'having heard the news he went out'), but it can also take on causal meaning, can describe simultaneous actions, or can be used to change an adjective into an adverb (e.g., ভাল /bhalo/ 'good' becomes ভাল করে /bhalo kore/ 'well'). It is not unusual to have a number of perfective participles in one sentence to describe consecutive events. Perfective participles are also used in the formation of compound verbs, in which two verbs combine to take on a new meaning. The second verb can lose its original meaning entirely and instead add an aspectual feature to the perfective participle, as in খাওয়া /khaoṣa/ *eat*, খেয়ে ফেলা /kheṣe phæla/ (lit: having eaten, throw = 'eat up') and আসা /asa/ 'come,' এসে পড়া /ese pōṛa/ (lit: having come, fall = 'arrive'). To some extent, nonfinite verb forms take over the role of subordinate clauses.

Impersonal structures are very common, as, for instance, in expressing possession, possibility, obligation, and physical sensations, feelings, and experiences (examples (3)–(6), respectively):

- (3) আমার গাড়ি আছে।
/amar gaṛi ache/
my car be.3.PERS PRES
'I have a car.'
- (4) এখানে যাওয়া যায়।
/ekhane yaoṣa yay/
there go.VN go.3.PERS PRES
'It is possible to go there.'
- (5) তাকে যেতে হবে।
/take yete hōbe/
him.ACC go.INF be.3.PERS FUT
'He will have to go.'

- (6) আমার ভয় লাগে।
 /amar bhoý lage/
 my fear attach.3.PERS PRES
 ‘I am afraid.’

The logical subject is usually in the genitive.

Passives are formed with verbal nouns and the verb হওয়া /hoýa/ ‘be, become’; for example, সে আমাকে টাকা দিয়েছে। /se amake ðaka diýeche/ ‘he has given me money’ becomes as shown in example (7):

- (7) আমাকে টাকা দেওয়া হয়েছে।
 /amake ðaka deoýa hoýeche/
 me. ACC money give.VN be.3.PERS PRF
 ‘The money has been given to me.’

Intransitive verbs can also be used in passive structures; for example, আমি যাবো। /ami yabo/ ‘I will go’ becomes as shown in example (8):

- (8) আমার যাওয়া হবে।
 /amar yaýoa hobe/
 my go.VN be.3.PERS FUT
 ‘My going will be.’

Special Features

If languages can be said to have particular characteristics, then Bengali has a sense of play in its phonetic structure. We find it in numerous onomatopoeia, such as চকচক /çokçok/ ‘glittering,’ টিপটিপ /tipip/ ‘dripping’ (water), ঘোঁত ঘোঁত /ghôṭghôṭ/ ‘grunting,’ খিলখিল /khil-khil/ ‘giggling,’ and ধু-ধু /dhu-dhu/ (expressing ‘desolation’), but also in sequences of similar or identical syllables to express mutual or extended actions, as in হাসাহাসি /hasahasi/ ‘laughing,’ মারামারি /maramari/

‘fighting,’ ঠেলাঠেলি /thælatheli/ ‘jostling,’ বকাবকি /boka-boki/ ‘bickering,’ লেখালেখি /lækhalekhi/ ‘correspondence,’ and কান্নাকাটি /kannakati/ ‘continuous weeping.’ Reduplication of adjectives and adverbs has an intensifying effect, as in বড় বড় /bôro bôro/ ‘big big’-‘very big,’ দূরে দূরে /dure dure/ ‘far far’ = ‘a long way away,’ and সকাল সকাল /sokal sokal/ ‘morning morning’-‘very early.’ Many of these combinations have an element of improvisation and greatly add to the charm of the language.

Bibliography

- Anderson J D (1920). *A manual of the Bengali language*. Cambridge: Cambridge University Press. [Reprinted in 1962.]
- Bykova E M (1962). ‘The Bengali language, questions on the grammar.’ In *USSR Academy of Sciences, languages of Asia and Africa*. Moscow: Nauka Publishing House. [Reprinted in 1981.]
- Chatterji S K (1926). *The origin and development of the Bengali language*. Calcutta: Calcutta University Press.
- Maniruzzaman (1986). ‘Linguistic studies on Bangla.’ In *Studies in the Bangla language*. Chittagong: University of Chittagong.
- Radice W (1994). *Teach yourself Bengali*. London: Hodder & Stoughton [2nd edition 2003].
- Singh U N (1986). *Bibliography of Bengali linguistics*. Mysore: Central Institute of Indian Languages.
- Smith W L (1997). *Bengali reference grammar*. Stockholm: Association of Oriental Studies.
- Thompson H R (1999). *Essential everyday Bengali*. Dhaka: Bangla Academy [2nd edition 2003].
- Zbavitel D (1970). *Non-finite verb forms in Bengali*. Prague: Czechoslovak Academy of Sciences.

Benue–Congo Languages

K Williamson

© 2006 Elsevier Ltd. All rights reserved.

The Benue–Congo languages form a very large group in Africa and include the well-known Bantu languages. The term ‘Benue–Congo’ was introduced by Greenberg (1963) to refer to one of the six branches of his Niger–Congo family. Previously, the Bantu languages had been treated as a separate family and the similarity of the other Benue–Congo languages to Bantu had been recognized by referring to them as ‘Semi–Bantu’ (Johnston, 1919–1922) or ‘Bantoid’ (e.g., Guthrie, 1948), equivalent to the ‘Benue–Cross’

of Westermann (1927). Greenberg’s innovation was to remove the separate status of Bantu, add it to Westermann’s Benue–Cross as a subgroup, and rename the group, using the term ‘Congo’ to indicate its extension into the Bantu area.

Greenberg’s View of Benue–Congo

Greenberg contrasted Benue–Congo with the other five branches of Niger–Congo, though he noted it was particularly close to Kwa. Internally, he subdivided it into Plateau, consisting of seven numbered subgroups; Jukunoid; Cross River, consisting of three numbered subgroups; and Bantoid, containing seven

languages or groups, the last of which is Bantu. The term ‘Old Benue–Congo’ refers to this scenario.

Views of Benue–Congo in the Late Twentieth Century

Bennett and Sterk (1977) noted that lexicostatistics led to some major changes in Greenberg’s scenario. In particular, they split Kwa in half and combined the Eastern half with Benue–Congo. The approximate consensus is presented in Bendor-Samuel (1989); Benue–Congo, including the former Eastern Kwa, is now one of the branches of Volta–Congo, which is in its turn a branch of Atlantic–Congo, within Niger–Congo. The term ‘New Benue–Congo’ refers to this scenario.

Subgrouping of (New) Benue–Congo

Because of the reclassification in the late 1980s and the very large number of languages involved, the subgrouping of New Benue–Congo is in a fluid state. On the basis of lexical innovations, Blench (1989) has suggested a major division between Western Benue–Congo, corresponding to the former Eastern Kwa, and Eastern Benue–Congo, corresponding to Old Benue–Congo. The recognized subgroups are now listed. (Nigerian orthographic conventions used in language names are as follows: ɔ [ɔ], ɛ [ɛ], ɪ [ɪ], ɔ̃ [ɔ̃], ɛ̃ [ɛ̃], ɛ̃ [ɛ̃].) Western Benue–Congo (formerly Eastern Kwa):

- (a) Ọkọ (Ogori): a small, little-studied language.
- (b) Ukaan–Akpes: two clusters of tiny, barely studied dialects.
- (c) Defoid: two clusters of tiny Akokoid (Amgbe) dialects, plus the Yoruboid group, comprising Yoruba, Işekiri, and Igala.
- (d) Edoid: a large number of languages, including Edo (Bini), and Urhobo.
- (e) Nupoid (Niger–Kaduna): some seventeen languages including Epira (Igbirra), Gade, Gbagyi and Gbari (jointly called Gwari), Kakanda, and Nupe.
- (f) Idomoid: some nine languages, including Idoma.
- (g) Igbooid: comprises Ẹkpẹye and a large language cluster centered around Igbo.

Eastern Benue–Congo (Old Benue–Congo):

- (h) Kainji: corresponds to Greenberg’s Plateau 1; subdivided into Western Kainji, including the Kambari and Bassa groups and the Lela (Dakarkari) language, and Eastern Kainji, including the Northern Jos group of small languages.
- (i) Platoid: corresponds to Greenberg’s Plateau 2–7 plus Jukunoid; subdivided into Plateau, with five geographical subgroups including many languages,

such as Eggon, Che (Rukuba), Berom, Jju (Kaje), and Tyap (Katab); and Benue, containing Tarok and related languages in one group and Jukunoid, including Jukun, in another.

- (j) Cross River: subdivided into Bendi, corresponding to Greenberg’s Cross River 1 and including Bekwarra and Bokyi; and Delta–Cross, corresponding to Greenberg’s Cross River 2 and 3 combined, comprising four subgroups: Upper Cross, including Mbembe and Lokaa, Lower Cross, including Anaang, Efik, Ibibio, and Obolo; Ogoni (Kegboid), including Kana, Gokana, and Eleme; and Central Delta, including Abuan and Ogbja.
- (k) Bantoid: subdivided into Northern Bantoid, comprising Mambila with related languages and Samba Daka with related languages; and Southern Bantoid, comprising the Bantu languages, taken in the broad sense, as used by Greenberg, with the addition of Tiv and languages related to it.

Geographical Location

The Benue–Congo language groups are chiefly found in Nigeria, with Yoruboid, Jukunoid, Cross River, and Northern Bantoid extending slightly into neighboring countries and Bantu having expanded dramatically into Central, East, and Southern Africa.

Typological Characteristics of the Group

Benue–Congo languages have Subject–Verb–Object or occasionally Subject–Modal–Verb–Object word order; adverbials are normally sentence-final. A wide variety of serial verb and consecutive verb constructions are found.

The most typical morphological feature is the existence of noun class systems, usually marked by paired singular/plural prefixes or, for mass nouns, by a single prefix. Words that qualify the noun show concord prefixes, and the verb also shows concord with the noun class of its subject. Some languages have developed noun class suffixes in addition to or instead of prefixes. Bantu languages are the most conservative in showing very full noun class systems, but there are few Benue–Congo languages that do not display at least remnants of a former noun class system.

Verbs often take suffixes, ‘verbal extensions’ or ‘extensional suffixes,’ which add such meanings as causative, reciprocal, or separative to the meaning of the root.

Most Benue–Congo language groups show typical phonological features of Niger–Congo, such as vowel harmony, labial-velar stops, and tone. The typical root structure is CVCV (where C = Consonant,

V = Vowel) in the more conservative languages; others have reduced their roots to CVC or CV. Complex nasal phenomena involving both vowels and consonants are widespread.

Bibliography

- Bendor-Samuel J (ed.) (1989). *The Niger–Congo languages*. Lanham, MD: University Press of America.
- Bennett P R & Sterk J P (1977). ‘South Central Niger–Congo: a reclassification.’ *Studies in African Linguistics* 8, 241–273.

- Blench R (1989). ‘New Benue–Congo: a definition and proposed internal classification.’ *Afrikanistische Arbeitspapiere* 17, 115–147.
- Greenberg J H (1963). *The languages of Africa*. The Hague: Mouton.
- Guthrie M (1948). *The classification of the Bantu languages*. London: Oxford University Press for International African Institute.
- Johnston H H (1919–1922). *A comparative study of the Bantu and Semi-Bantu languages*. Oxford: Clarendon Press.
- Westermann D (1927). *Die westlichen Sudansprachen und ihre Beziehungen zum Bantu*. Berlin: De Gruyter.

Berber

S Chaker, INALCO, Paris, France

A Mettouchi, University of Nantes, Nantes, France

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Berber language is one of the branches of the large Hamito-Semitic (Afroasiatic) linguistic family, which also includes Semitic, Cushitic, ancient Egyptian, and Chadic. With all that this notion implies, Berber can be considered as the ‘aboriginal’ language of North Africa because currently there is no positive trace of an exterior origin or of the presence of a pre- or non-Berber substratum in this region. As far back as one can go (first Egyptian accounts: cf. Bates, 1914/1970), the Berber language was already installed in its present territory. Particularly, the toponymy has not allowed us to identify, up till now, any kind of pre-Berber linguistic sediment. Despite numerous theories suggested by linguists since the 19th century in favour of an external origin of the language (Middle East or East Africa), neither prehistoric archaeology nor physical anthropology could show the movement of a population coming from elsewhere; it has even been solidly established that man has been present in North Africa, in a continuous manner, for at least a million years (cf. Camps, 1974, 1980).

Tamazight (the Berber word for language) covers a vast geographical area: all of North Africa, the Sahara, and a part of the West African Sahel. But the countries principally concerned are, by order of demographical importance: Morocco (35–40% of the total population), Algeria (25% of the population), Niger, and Mali (Tuaregs) (Figure 1).

The Berber-Speaking Regions

In Morocco, spoken Berber is spread into three large dialectal areas that cover the totality of the mountainous regions: in the north is the Rif (Tarifit); in the center, the Mid-Atlas and a part of the High-Atlas (Tamazight [Tamazight, Central Atlas]); and in the south/southwest (High-Atlas, Anti-Atlas and Under), the Chleuh domain (Tachelhit/Tašelhit/Chilha).

In Algeria, the principal Berber-speaking region is Kabylia. In a relatively limited but densely populated surface area, Kabylia (Kabyle; Taqbaylit dialect) alone has two-thirds of Algeria’s Berber speakers. The other significant Berber-speaking groups are: the Chaouias (Chaouia; Tachawit) of the Aures region, having in all likelihood a million people, and the people of the Mزاب (in Ghardaia and other Ibadhite cities), having a population of between 150 000 and 200 000. There are in fact other Berber-speaking groups in Algeria, but these are modest linguistic islands of only several thousands to tens of thousands of speakers.

The third large group of Berber speakers is the Tuaregs (Tamashaq [Tamasheq], Tamajaq [Tamajaq, Tawallammat], Tamahaq [Tamahaq, Tahaggart]), straddling several countries across the Sahara-Sahel zone, principally in Niger (±500 000 people) and in Mali (450 000). The other countries: Algeria (Ahaggar, Ajjer dialects), Libya (Ajjer dialect) Burkina-Faso, and even Nigeria, have more limited Tuareg populations. The total Tuareg population is well over 1 million individuals.

The other Berber speaking regions are isolated, often threatened areas, spread out across the south of Mauritania (Zenaga), in Tunisia (in Djerba, in part, and in several villages in the south-central part

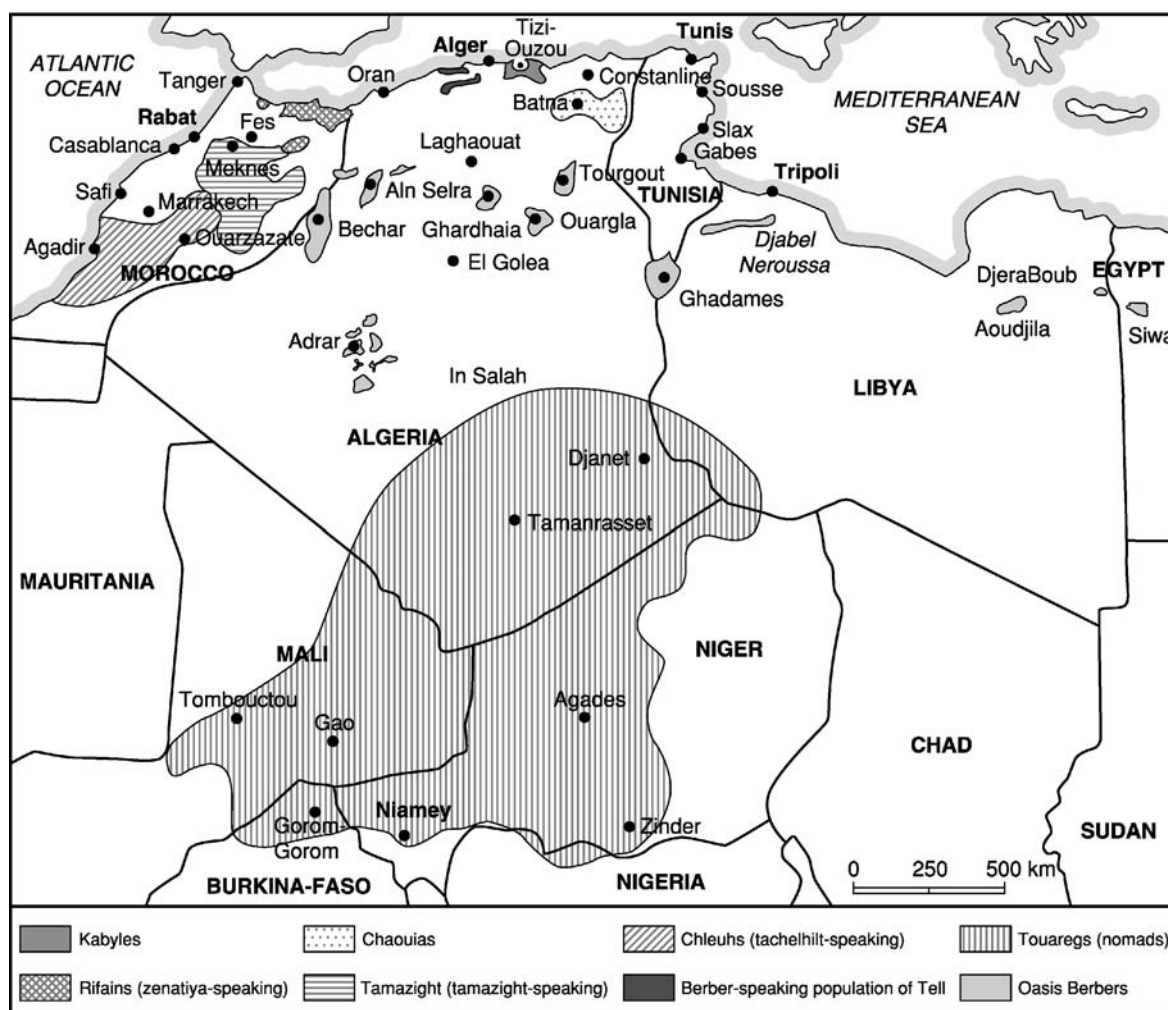


Figure 1 Map of the Berber-speaking region in North Africa.

of the country), in Libya (where Berber-speaking groups are clearly larger and more resistant), and in Egypt (the Siwa Oasis).

But these are only the traditional locations: from the beginning of the 20th century and especially since decolonization, worker emigration and the massive rural exodus that took place throughout the Maghrib have been the basis for the formation of Berber-speaking communities in all the major cities: Algiers and Casablanca are the most outstanding examples. And Paris is one of the three principal Berber-speaking cities of the world, perhaps even the largest!

Linguistic Features

Phonetics and Phonology

The phonological consonantic system of Berber (Basset, 1952/1969; Prasse, 1972–1974) relies on an opposition between tensed and nontensed consonants. Variation is induced by: phonemes borrowed from

Arabic (Arabic, Standard) (pharyngeals, some emphatics), a tendency towards spirantization in Northern dialects, and palatalization and labio-velarization.

The vocalic system of Berber is ternary: /a/ vs. /i/ vs. /u/. The schwa [ə] is considered by most researchers as a neutral vowel without phonological status. Intermediary phonemes (/e/, /o/, /ä/) that exist in some dialects (Tuareg, Libya, Tunisia) are recent innovations (Prasse, 1984–1986), stemming from the probable phonologization of former contextually conditioned variants. The same is also probably true of vocalic duration, which has distinctive status in those dialects (for instance, to mark the intensive perfective in Tuareg). It probably originates in an expressive lengthening, or in a quantitative reinterpretation of accentual phenomena.

Morphology

Berber stems are composed of a consonantal root and an inflectional scheme, which is specific to the

considered part of speech. There are, for instance, adjectival schemes, verbal (aspectual) schemes, and nominal schemes (Table 1, Table 2).

The verb ‘go/go with’, is composed of a root (*dd*) and an obligatory aspectual inflection (Table 3).

The morphology of Berber is heavily derivational. For instance, there is a class of labile (ambitransitive) verbs, which varies in size depending on the dialects, and which can be semantically transitivized thanks to a causative-transitive prefix (*s-*). The ‘passive’ is rare and marked through a *ttw-* prefix, reciprocals and middles are marked thanks to a nasal prefix (*my-*). Those prefixes have variants within each dialect (Table 4).

Case in Berber is limited to an opposition between what is traditionally called ‘état libre’ and ‘état d’annexion.’ The former is unmarked, and the latter marked. ‘État libre’ is the form taken by nominals in citation form, topic position, direct object, possessee.

Table 1 Adjectival scheme

| Adjective | Verb | Root | Adj. Scheme | Adjective |
|-----------|--|--------------------------|---------------|----------------|
| ‘white’ | <i>i-mlul</i> sub.3MSG- be.white | <i>mll</i> <i>ccc</i> | <i>acc:ac</i> | <i>aməllal</i> |

Table 2 Nominal scheme

| Noun | Verb | Root | Nom. Scheme | Adjective |
|----------|--|------------------------|----------------------------------|---------------------------|
| ‘robber’ | <i>y-uk^wər</i> sub.3MSG- be.white | <i>kr</i> <i>cc</i> | (agent noun) <i>am-vcc</i> | <i>amak^war</i> |

Table 3 Aspectual inflections (Taqbaylit)

| Verb | Root | Aorist | Perfective | Negative Perfective | Imperfective |
|------|-----------|------------|------------|---------------------|--------------|
| ‘go’ | <i>dd</i> | <i>ddu</i> | <i>dda</i> | <i>ddi</i> | <i>təddu</i> |

Table 4 Verbal derivation

| Stem | Prefix | Verb | Grammar | Gloss |
|-------------|---------------|------------------|---------------------------|----------------------------|
| <i>kkəs</i> | + <i>s-</i> | <i>su-kkəs</i> | CAUS-take.off.PERFECTIVE | ‘made X take off’ |
| ‘take off’ | + <i>ttw-</i> | <i>ttwa-kkəs</i> | PASS-take.off.PERFECTIVE | ‘got taken off’ |
| | + <i>my-</i> | <i>my-kkəs</i> | RECIP-take.off.PERFECTIVE | ‘took off from each other’ |

‘État d’annexion’ is the form taken by postverbal subjects, nominals following prepositions and numerals, possessors (Table 5).

This distinction is no longer alive in all dialects. Dialects that have lost the opposition are: Nefoussa, Ghadames, Sokna, Siwa (Siwi) in Egypt and Zenaga of Mauritania.

There are two genders, masculine (unmarked), and feminine (marked). Gender is arbitrary. Feminine gender can function as a diminutive or partitive, or denote an item as opposed to a collection (Table 6).

Number distinctions are between singular and plural. Plural inflections are varied, either formed by affixation, or apophony; some plurals are irregular (Table 7).

There are no articles in Berber. Definiteness is contextually inferable, word order playing a role in the matter. Anaphoric and deictic particles appear where necessary to disambiguate.

All verbs are completed with a personal or participial affix. Therefore, the minimal utterance is composed of a root, always inflected for aspect, and its obligatory personal affix (and accusative and dative clitics where applicable):

- (1) *yə-čča* (Taqbaylit)
SUB.3MSG-eat.PERFECTIVE
he ate/has eaten
- (2) *yə-fka* *yas*
SUB.3MPL-give.PERFECTIVE DAT.3SG
t *idd*
ACC.3MSG proximal.particle
He gave it to her/him

Constituent Order

Such minimal utterances are very frequent in authentic speech. However, longer utterances, containing noun phrases, also appear. The maximal configuration is exemplified below, and illustrates the VSO type:

- (3) *yə-fka* *umɣar* *idrimən*
SUB.3MSG-give.PERFECTIVE old.man.EA money.EL
i *umddak^wəl-is*
to comrade.EA-POSS.3MSG
the old man gave (some) money to his
companion

Table 5 Case

| | <i>État libre (EL)</i> | <i>État d'annexion (EA)</i> |
|--------|---|---|
| clause | <i>tafunast tə-čča</i> COW.EL SUB.3FSG- eat.PERFECTIVE (As for) the cow (she) ate/ has eaten. | <i>tə-čča tafunast</i> SUB.3FSG-eat.PERFECTIVE cow.EA The cow ate/has eaten. |
| phrase | <i>axxam uməkxa</i> house.EL shepherd.EA The shepherd's house | <i>axxam uməkxa</i> house.EL shepherd.EA The shepherd's house |

Table 6 Gender

| <i>Noun Form</i> | <i>Masculine</i> | <i>Feminine</i> |
|--------------------------|-------------------------|---------------------------------------|
| diminutive/ partitive | <i>axxam</i> 'house' | <i>taxxamt</i> 'small house/ room' |
| collective vs. item/ | <i>azəmmur</i> 'olives' | <i>tazəmmurt</i> 'olive tree' |

Table 7 Number

| <i>Number</i> | <i>'house'</i> | <i>'braid'</i> | <i>'heart'</i> | <i>'town'</i> |
|---------------|----------------|----------------|----------------|-----------------|
| singular | <i>axxam</i> | <i>asaru</i> | <i>ul</i> | <i>tamdint</i> |
| plural | <i>ixxamən</i> | <i>isura</i> | <i>ulawən</i> | <i>timdinin</i> |

Only a few quantitative studies on word order have been conducted. Among them, Mettouchi (to appear a) showed that in Taqbaylit, word order was in fact pragmatically motivated. This motivation is also probably true of other dialects. The following table (where V actually stands for a minimal utterance (root+personal affix)) shows the various configurations encountered in authentic speech (Table 8).

This table shows that, whereas the characterization of Berber (here Taqbaylit) as a VO language seems to hold, the status and position of the 'subject' is somewhat more problematic: almost one-fourth of the utterances can appear without one. This special behavior of the subject in Berber has long been recognized in Berber studies. Thus, traditionally, it is the personal affix that is considered as the real subject (and not as an agreement marker), whereas the preverbal coreferential nominal is called 'indicateur de thème' and the postverbal coreferential nominal 'complément explicatif' (Galand, 1964/2002). The positions of nominal constituents are determined to a large extent by pragmatic and semantic factors. Taqbaylit can therefore be considered as a nonconfigurational language, and more precisely, as a pronominal argument language. Quantitative studies

Table 8 Constituent order found in a conversational excerpt (143 third-person verbal predications)

| VS | SV | V | OV | VO |
|-----------------------------|-------------------------------|-------------|-------------|-------------|
| 60 42% (incl. VSO 3%) | 25 17.5% (incl. SVO 2%) | 35 24.5% | 1 0.5% | 22 15.5% |
| | 85 59.5% | | 58 40.5% | |

Table 9 Participial circumfixes

| <i>Aïr</i> | <i>Singular</i> | <i>Plural</i> | <i>Taqbaylit</i> | <i>Singular</i> | <i>Plural</i> |
|---------------|-----------------|---------------|------------------|-----------------|---------------|
| <i>Tuareg</i> | | | | | |
| masc. | y—n | | masc. | y— | -n |
| | | nin | | | |
| fem. | t—t | | fem. | | |

must be conducted on other dialects to see whether this characterization is valid for Berber as a whole.

Berber is head marking at the level of the clause, but dependent marking at the level of the phrase. At the level of the phrase, Berber is also more rigid, and has the following properties among Greenberg's universals: it has prepositions, the possessor follows the possessee, the modifier (as well as relative clauses) follows the head noun and affixes are mostly prefixes.

Relative clauses (Galand, 1988) are distinguished according to the status of the antecedent: if it is coreferential to the subject of the relative clause, a participle is used. This form is composed of a root inflected for aspect, and an invariant circumfix (in Taqbaylit), or a limited set of affixes (in Tuareg) (Table 9).

In some syntactic contexts (relative clauses, interrogation, negation, TAM preverbs), clitics change position and attach themselves to the new head of the sentence (negative marker, interrogative pronouns or relativizer, preverb). This phenomenon of clitic-climbing is exemplified below:

- (4) ad as t idd
irrealis DAT.3SG ACC.3MSG *proximal.particle*
 yə-fk
 SUB.3MSG-give.PERFECTIVE
 He will give it to her/him

Predicate Nominals and Related Constructions

Verbs very often are the center of predication, but predicates can also be nonverbal. Nouns, adjectives, and free pronouns can function as predicates.

Attribution is marked, either, as in Tuareg, through a simple juxtaposition of nouns

- (5) Mûsa amɣar n Ahaggar
Mûsa chief.EL of Ahaggar
 Mûsa is the chief of Ahaggar

or, as in most Northern Berber dialects, thanks to a special invariant copula (particle):

- (6) d amur-iw
predicative.particle share.EL-my
 It's my share

Focus constructions are mostly based on attributive clauses (Taqbaylit):

- (7) d amur-iw i
predicative.particle share.EL-my relator
 dd y-uk^wər
prox.particle SUB.3MSG-steal.PERFECTIVE
 It's my share that he stole

But focus-fronting (traditionally called 'anticipation renforcée') is also encountered (Tuareg):

- (8) tagəlla a tə-kša
bread DEMONSTRATIVE SUB.3FSG-eat.PERFECTIVE
 təmɣart
old-woman.EA
 It's bread that the old woman has eaten

Attributive predication can also be expressed thanks to a special category of verbs, quality verbs, which are only alive in some dialects, among them Taqbaylit. This category represents approximately 60 verbs, mostly referring to size and color, but also to other, more unexpected, semantic domains (Chaker, 1983: 117–118). It is characterized morphologically by a special suffixal conjugation in the perfective, in the 3rd person and the plural. Here is the paradigm of affixes for the verb *məqq^wr* 'be big' (Table 10).

Table 10 Quality verbs

| Person | Singular | Plural |
|--------|----------------------------|-----------------------------|
| 1 | <i>məqq^wr-ɣ</i> | |
| 2 | <i>məqq^wr-d</i> | <i>məqq^wr-it</i> |
| 3 M | <i>məqq^wr</i> | |
| 3 F | <i>məqq^wr-t</i> | |

'Existence' is marked thanks to the verb *ili* 'be,' 'exist,' in the perfective (Taqbaylit).

- (9) lla-n waman
exist.PERFECTIVE-SUB.3MPL waters.EA
 There is water

Location can be predicated thanks to the association of an interrogative pronoun and an accusative clitic (Taqbaylit):

- (10) anda t umur-iw?
where ACC.3MSG share.EA-my?
 where is my share?

Possession is mostly predicated through the association of a preposition and a special personal affix (Taqbaylit):

- (11) ɣur-s sin yəzgarən
with-him two oxen.EA
 He has two oxen

Aspect

Berber dialects are basically aspectual, with evolutions towards tensedness in some of them (Tachelhit, cf. Leguil, 1992). A. Basset (1929, 1952/1969) was the first to reconstruct the basic ternary system of Berber, which opposes three forms: aorist ('aoriste simple'), perfective ('accompli,' 'prétérit'), and imperfective ('inaccompli,' 'aoriste intensif') (Table 11).

All dialects have a special negative form (negative perfective, called 'accompli négatif' or 'prétérit négatif') that is used instead of the perfective after the negative marker. Some dialects also have secondary, more recent, forms: negative imperfective ('inaccompli négatif'), and resultative perfective ('accompli résultatif'). Here is for instance the full system of Tuareg (Table 12).

In all dialects, those forms are preverbed by TAM markers, giving rise to various configurations. Taking into account the preverbs is absolutely necessary to describe properly the oppositions at stake in Berber (Chaker, 1997). Among those preverbs, the most frequent cross-dialectally are *ad* (irrealis), *rad* (future), and *la* (progressive). They stem from ancient deictic or locative markers, and from auxiliaries.

Moreover, verbal negation (*ur*) acts on those oppositions, giving rise to asymmetries (Mettouchi, to

Table 11 Basic aspectual opposition

| Aorist | Perfective | Imperfective |
|--|---|--|
| <i>y-ak^wər</i> SUB.3MSG-steal.aorist neutral/indefinite | <i>y-ukər</i> SUB.3MSG-steal.perfective punctual/definite/completed | <i>yə-ttak^wər</i> SUB.3MSG-steal.imperfective durative/iterative/habitual/progressive |

Table 12 Aïr Tuareg aspectual bases

| Roots | Aorist | Perfective | Perfect | Neg. Perfective | Imperfective | Neg. Imperfective |
|---------------------|-----------------|-------------------|-------------------|-------------------|---------------------|--------------------|
| <i>rtk</i> , 'fall' | - <i>rtək</i> - | - <i>rtak</i> - | - <i>rtaak</i> - | - <i>rtek</i> - | - <i>raattək</i> - | - <i>rettək</i> - |
| <i>g</i> , 'do' | - <i>g(u)</i> - | - <i>ge/a/ə</i> - | - <i>gee/aa</i> - | - <i>ge/a/ə</i> - | - <i>taagg(u)</i> - | - <i>tagg(u)</i> - |

Table 13 Negation and aspectual asymmetry in Taqbaylit^a

| Positive | Negative |
|--|---|
| aorist (optative, imperative) 1% | <i>a wər</i> + aorist (optative) <1% |
| <i>ad</i> + aorist 30% | <i>ur</i> + imperfective 37% |
| <i>la/ad/Ø</i> + imperfective 16% | |
| perfective 53% | <i>ur</i> + negative perfective 63% |
| 100% of positive utterances | 100% of negative utterances |

^aFrequency counts are based on a conversational corpus.

appear b). Here are, for instance, the actual oppositions encountered in Taqbaylit (Table 13).

Further Resources

A systematic, regular bibliographic orientation can be found in the *Annuaire de l'Afrique du Nord* (Paris, CNRS) since 1965 (volume IV), edited by Lionel Galand, then Salem Chaker and Claude Brenier-Estrine. There is also a recent, very complete bibliographic recapitulation in *Langues et littératures berbères des origines à nos jours*. Bibliographie internationale (Paris, Ibis Press, 1997), and a bibliographic database developed by Salem Chaker, that can be queried online on the Internet site of the Berber Research Center.

Bibliography

- Abdelmassih E T (1968). *Tamazight verb structure: a generative approach*. Bloomington: Indiana University/The Hague: Mouton.
- Applegate J R (1970). 'The Berber Languages.' In *Current trends in linguistics* 6. Paris/La Haye.
- Basset A (1929). *La langue berbère. Morphologie. Le verbe. Etude de thèmes*. Paris.
- Basset A (1952/1969). *La langue berbère*. London: International African Institute.
- Bates O (1914/1970). *Eastern Libyan*. London: Cass.
- Camps G (1974). *Les civilisations préhistoriques de l'Afrique du Nord et du Sahara*. Paris: Doin.
- Camps G (1980). *Berbères: aux marges de l'histoire*. Toulouse: Editions des Hespérides. Re-published under the title: *Berbères. Mémoire et identité*, Paris, Editions Errances, 1987.

- Chaker S (1977/1981). 'Une inscription libyque du Musée des Antiquités d'Alger.' *Libyca* 25, 193–202.
- Chaker S (1983). *Un parler berbère d'Algérie (Kabylie): syntaxe*. Université de Provence.
- Chaker S (1984). *Introduction au domaine berbère: Textes en linguistique berbère*. Paris: CNRS.
- Chaker S (1995). *Linguistique berbère: études de syntaxe et de diachronie*. Paris/Louvain, Editions Peeters.
- Chaker S (1997). 'Quelques faits de grammaticalisation dans le système verbal berbère.' In *Grammaticalisation et reconstruction: Mémoires de la Société de Linguistique de Paris*, n. s. 103–121.
- Chaker S (1998). *Berbères aujourd'hui* (2nd edn.). Paris: L'Harmattan.
- Galand L (1957). 'Un cas particulier de phrase non-verbale: 'l'anticipation renforcée' et l'interrogation en berbère.' In *Mémorial André Basset*. Paris: Maisonneuve. 27–37.
- Galand L (1964). 'L'énoncé verbal en berbère.' *Cahiers Ferdinand de Saussure* 21, 33–53, reprinted in Galand L (2002). *Etudes de Linguistique Berbère*, pp. 165–175.
- Galand L (1977). 'Continuité et renouvellement d'un système verbal: le cas du berbère.' *Bulletin de la Société de Linguistique de Paris* 72(1).
- Galand L (1988). 'Typologie des propositions relatives: la place du berbère.' *LALIES* 6, 81–101.
- Galand L (1989a). 'Les alphabets libyques.' *Antiquités africaines* 25, 69–81.
- Galand L (1989b). 'Les langues berbères' *La réforme des langues: Histoire et avenir IV*. Hamburg: H. Buske Verlag.
- Galand L (2002). *Etudes de linguistique berbère*. Paris/Louvain: Peeters (Publications de la Société de Linguistique de Paris).
- Greenberg J (1966). *Languages of Africa*. The Hague: Mouton.
- Leguil A (1992). *Bilan et perspectives: Structures prédictives en berbère*. Paris: L'Harmattan.
- Mettouchi A (2003). 'Focalisation contrastive et structure de l'information en kabyle (berbère).' In François J & Lacheret A (eds.) *Mémoires de la Société de Linguistique de Paris: Fonctions et moyens d'expression de la focalisation*. Peeters, Paris/Louvain. 81–97.
- Mettouchi A (to appear a). 'Discourse-configurationality and the encoding of semantic macroroles in Taqbaylit Berber: noun phrases, personal affixes and clitics.' In Mengozzi A (ed.) *Proceedings of the XI Incontro Italiano di linguistica Camito-Semitica*. Bergamo, 5–7, June 2003.
- Mettouchi A (to appear b). 'The System of Negation in Berber.' Cyffer N & Ebermann E (ed.) *Typological Studies in Language: Patterns of negation in West African languages*. Amsterdam: J. Benjamins.

- Penchoen T G (1973a). *Studi Magrebini V: Etude syntaxique d'un parler berbère (Aït Frah de l'Aurès)*. Napoli.
- Penchoen T G (1973b). *Tamazight of the Ayt Ndhir*. Los Angeles: Undena Publications.
- Prasse K G (1972–1974). *Manuel de grammaire touarègue (tahaggart) I–III: Phonétique-Ecriture-Pronom; IV–V: Nom; VI–VIII: Verbe*. Copenhagen: Akademisk, Forlag.
- Prasse K G (1984–1986). 'The values of the tenses in Tuarég (Berber).' *Orientalia Suecana* 33–35, 333–339.

- Salama P (1993). 'A propos d'une inscription libyque du Musée des Antiquités d'Alger.' In *GLECS 15: A la croisée des études libyco-berbères: melange offerts à Paulette Galand-Pernet et Lionel Galand*. Paris: Geuthner. 127–140.

Relevant Website

<http://194.167.236.5/pub/enseignements/langues/afrique/berbere> – Berber Research Center.

Bikol

J W Lobel, University of Hawai'i, Manoa, Honolulu, HI, USA

© 2006 Elsevier Ltd. All rights reserved.

Bikol refers to a group of three related Austronesian languages spoken in the Bikol Region of the southern Luzon peninsula of the Philippines, one of the central Philippines' most dialectally diverse areas. As a branch of the Central Philippine subgroup, Bikol is coordinate with the Tagalog and Bisayan branches. The Bikol branch contains 2 of the 12 major Philippine languages (i.e., those having more than 1 million speakers), with Northern Bikol (Central Bicolano) (including the standard Bikol of the cities of Naga and Legaspi) having 2.5 million speakers, and Southern Bikol (Bicolano, Albay), with about 1 million speakers. A third language, Northern Catanduanes Bikol (Bicolano, Northern Catanduanes), has a population of approximately 100 000. Most speakers of a Bikol language simply refer to their language as 'Bikol' without any further distinction of the specific language or dialect they speak. Furthermore, the name Bikol is also used by native speakers to refer to two dialects in central and southern Sorsogon province, even though these dialects are generally classified as Central Bisayan dialects with heavy lexical borrowing from Bikol.

All of the Bikol languages are spoken natively solely within the Bikol Region, a political unit that includes the six provinces of Camarines Norte, Camarines Sur, Albay, Sorsogon, Catanduanes, and Masbate. The Northern Bikol language consists mainly of dialects spoken in and around the major centers of Naga, Legaspi, Daet, and Virac, along with the entire northern coast of the Bikol peninsula from Vinzons in Camarines Norte to Prieto Diaz in Sorsogon, as well as in the town of San Pascual in Masbate Province, and Magallanes in central

Sorsogon. The Southern Bikol language consists of the Rinconada and Buhi-non dialects of southeastern Camarines Sur, the Miraya dialects in southwestern Albay and Donsol town in northwestern Sorsogon province, and the dialect of Libon, Albay. The Northern Catanduanes language is spoken in the northern half of the island of Catanduanes.

The standard dialect of the Bikol Region is the dialect of the cities of Naga and Legaspi, referred to as 'Bikol Naga' in the towns closer to Naga and as 'Bikol Legaspi' in the towns closer to Legaspi. The origin of this dialect's status can be traced to the end of the 16th century when Naga (formerly Nueva Caceres) was one of only three officially designated *ciudades* and the seat of one of only three bishops outside of Manila. (Doeppers, 1972) This dialect is still used by the church throughout the Bikol Region to the exclusion of all other varieties of Bikol. With the exception of Bikol Naga, most of the speech varieties of the Bikol Region are underdocumented. The only works that have been published on other dialects are a short description (Yamada, 1972) and textbook (Portugal, 2000) for Buhi-non, and a phrasebook (Lobel and Bucad, 2001b) for Rinconada.

The Bikol language (Naga dialect) was first documented by Marcos de Lisboa (d. 1622), whose *Vocabulario de la lengua Bicol* was published posthumously in 1754 and republished in 1865. Lisboa's work was preceded in print by Andres de San Agustin's (d. 1649) *Arte de la lengua Bicol*, first published in 1647, and republished in 1739, 1795, and 1879. Together, these two works represent the basis of nearly everything written about the Bikol language prior to the 20th century.

The major modern works on Bikol include a textbook (Mintz, 1971a), a grammar description (Mintz, 1971b, 1973), a dictionary (Mintz and del Rosario Britanico, 1985), and two descriptions of dialectology (McFarland, 1974 and Lobel and Tria, 2000).

Table 1 Standard Bikol pronouns

| <i>Person and number</i> | <i>Nominative case</i> | <i>Genitive case</i> | <i>Oblique case</i> |
|-----------------------------------|------------------------|------------------------------|---------------------|
| 1st singular | akô | ko ² (OBik nyákò) | sakô, sakúyà |
| 2nd singular | iká, ka ^b | mo (OBik nímò) | sáimo |
| 3rd singular | siyá | niyá | saíya |
| 1st exclusive ^a plural | kami | mi, nyámò | samô, samúyà |
| 1st inclusive ^a plural | kitá | ta, nyátò | satô, satúyà |
| 2nd plural | kamó | nindó | saindó |
| 3rd plural | sindá | nindá | saindá |

^aInclusive pronouns include the addressee, while exclusive pronouns do not.

^bThe portmanteau pronoun *taká* replaces the ungrammatical sequence *ko ká.

Table 2 Standard Bikol case markers

| <i>Case</i> | <i>Reference</i> | <i>Bikol Naga marker</i> | <i>Bikol Legaspi marker</i> |
|-------------|------------------|--------------------------|-----------------------------|
| Nominative | – referential | an | an |
| | + referential | si | su |
| Genitive | – referential | nin | ki |
| | + referential | kan | kan |
| Oblique | | sa | sa |

Table 3 Standard Bikol demonstratives

| <i>Location of person/object</i> | <i>Nominative</i> | <i>Genitive</i> | <i>Oblique</i> | <i>Locational</i> |
|-------------------------------------|----------------------|------------------------|------------------------|---|
| Near speaker | iní 'this' | kainí 'this' | digdí, igdí 'here' | yaon digdí (Bikol Naga) anion digdí (Bikol Legaspi) 'is here' |
| Near addressee, far from speaker | iyán, an 'that' | kaiyán, kan 'that' | diyán 'there' | yaon diyán (Bikol Naga) uya diyán (Bikol Legaspi) 'is there' |
| Far from both speaker and addressee | idtó 'that (far)' | kaidtó 'that (far)' | dumán 'there (far)' | yaon dumán (Bikol Naga) idtoón dumán (Bikol Legaspi) 'is there (far)' |

Table 4 Standard Bikol focus-mood-aspect morphology

| <i>Mood</i> | <i>Aspect</i> | <i>Actor focus</i> | <i>Object focus (1)</i> | <i>Object focus (2)/Beneficiary focus</i> | <i>Location focus</i> |
|------------------------|---------------------|--------------------|-------------------------|---|-----------------------|
| Indicative | infinitive | mag- | -on | i- | -an |
| | past/perfective | nag- | pig- | (i)pig- | pig-...-an |
| | | | -in- | i-...-in- | -in-...-an |
| | present/progressive | nag-R- | pig-R- | (i)pig-R- | pig-R-...-an |
| | | | -in-R- | -in-R- | -in-R-...-an |
| | future | mā- | R-...-n | i-R- | R-...-an |
| Abilitative/Accidental | imperative | -um-/ø- | -a | -an | -i |
| | negative | mag- | pag-...-on | ipag- | pag-...-an |
| | negative imperative | pag- | pag-...-a | pag-...-an | pag-...-i |
| | infinitive | maka- | ma- | ika- | ma-...-an |
| | past/perfective | naka- | na- | ikina- | na-...-an |
| | present/progressive | nakaka- | na-R- | ikinaka- | na-R-...-an |
| | future | makaka- | ma-R- | ikaka- | ma-R-...-an |

With the exception of the dialectology studies, all of these works concentrate exclusively on the standard Bikol of Naga or Legaspi.

During the first part of the 20th century, the Bikol Region was home to a relatively bustling literary

scene, but today there are only scattered efforts at reviving a written tradition, and very little can be found in print in the Bikol language other than the Bible and other religion-related materials. In the past decade, there have been efforts to introduce Bikol

language subjects in the schools, and several universities have offered electives in Bikol Language and Literature in recent years. In general, the local variety of Bikol is still the language of most daily transactions, with Tagalog and English being confined to educational institutions, most forms of media, and higher-level business and government transactions.

Bikol has the basic Central Philippine-type phonology with 16 consonants /p b m w t d n s l r y k ŋ ʔ h/, three vowels /a i u/, contrastive stress, and contrastive length. Some dialects of Southern Bikol have preserved a fourth vowel as a reflex of PAN *e, usually realized as a high, central tense vowel /i/ but also realized as /o/ in Libon. Two dialects have an extra consonant phoneme: Southern Catanduanes, which has an interdental lateral, and Buhi-non, which has a voiced velar fricative. The Bikol orthography is largely phonemic except that it does not represent stress, length, or the glottal stop.

Bikol is agglutinative, with a complex system of verbal morphology expressing a wide variety of semantic and syntactic contrasts. Although sometimes analyzed as ergative, these languages are probably of a separate type called Symmetrical Voice Languages in which multiple voice distinctions exist, yet none can be considered more basic than the other (Himmelman, to appear). Like most other Philippine languages, there are four main verbal voices or 'focuses' (Actor, Object, Location, and Beneficiary) and three case distinctions (Nominative, Genitive, and Oblique) marked on Noun Phrases, name phrases, and pronouns by an introductory morpheme. Nouns, adjectives, and verbs distinguish between singular, plural, and in some cases, dual, and verbs may also be marked for reciprocal action. A number of other meanings can be marked by verbal affixes, including accidental, abilitative, distributive, repetitive, causative, social, diminutive, and infrequentive. Tense-aspect-mood distinctions include infinitive, past/perfective, present/progressive, future, imperative, negative, and negative imperative. Both reduplication and repetition are productive mechanisms that can denote diminutive, repetitive, and intensive meanings, among others. Refer to tables 1–4 for more information about: pronouns, case markers, demonstratives, and focus-mood-aspect morphology.

The Bikol languages have much the same grammatical structure as Tagalog, except for (a) a preference for inflecting verbs for plural actors (with the infix -Vr-), (b) the existence of distinct imperative forms, (c) the indication of repetitive action by a verbal affix (-para-), and (d) a more elaborate system of case markers that distinguish between referential and nonreferential, and in some dialects, past vs. nonpast.

A noteworthy feature of the Bikol languages is the presence of a speech register reserved for use in anger (Mintz, 1991, Lobel to appear). The lexicon of this angry register is usually either loosely derived or totally unrelated to their normal, nonangry equivalents. As such, an utterance by an angry speaker may hardly resemble an utterance with the same meaning spoken by a nonangry speaker.

Bibliography

- Carpio C M (1974). A study of Bikol metrical romances. M.A. thesis: Quezon City: University of the Philippines.
- Cordero K (2004). *Mga Tulang Tulala: Piling Tula sa Filipino, Bikol at Rinconada*. Naga City.
- Doeppers D F (1972). 'The development of Philippine cities before 1900.' *Journal of Asian Studies* 31(4), 769–792. Ann Arbor: Association for Asian Studies.
- Dy-Liacco L R (1956). *Bikol syntax*. M.A. thesis: Naga.
- Dy-Liacco L R (1996). *Sarung Dolot Sa Satuyang Ina: an offering to our mother on the quadricentennial celebration of Christianity in Bikol*. Parañaque: J & R Printing Co.
- Espinas M B (1983). *Bikol voices anthology*. Legaspi City: Bikol University.
- Espinas M B (1996). *Ibalong: the Bikol folk epic-fragment: English and Bikol translations, views and comments*. Legaspi City: M. C. Espinas.
- Gerona D M (1988). *From epic to history: a brief introduction to Bicol history*. Naga City: Ateneo de Naga University.
- Himmelman N (to appear). 'Typological characteristics: The Austronesian languages of Asia and Madagascar.' In Himmelman N & Adelaar K A (eds.). London: Curzon Press.
- Lisboa M de (1865). *Vocabulario de la lengua Bicol*. Manila: Establecimiento Tipográfico del Colegio de Santo Tomás.
- Lobel J W (2004). 'Old Bikol -um- vs mag- and the loss of a morphological paradigm.' *Oceanic Linguistics* 43(2), 340–368.
- Lobel J W (to appear). *The angry register of the Bikol languages of the Philippines*.
- Lobel J W & Bucad G U (2000). *Bikol phrasebook: a tourist's guide to the Bikol language*. Naga City: Lobel & Tria Partnership, Co.
- Lobel J W & Bucad G U (2001a). *Bikol literature anthology 1*. Naga City: Lobel & Tria Partnership, Co.
- Lobel J W & Bucad G U (2001b). *Rinconada-Bikol-Filipino-English phrasebook with mini-dictionary*. Naga City: Lobel & Tria Partnership, Co.
- Lobel J W & Lobel G B (2004). *Bikol literature anthology 2*. Naga City: Lobel & Tria Partnership, Co.
- Lobel J W & Tria W J S (2000). *An Satuyang Tataramon: a study of the Bikol language*. Naga City: Lobel & Tria Partnership, Co.
- McFarland C D (1974). The dialects of the Bikol area. Ph.D. diss.: Yale University.

- Mintz M W (1971a). *Bikol text*. Honolulu: University of Hawaii Press.
- Mintz M W (1971b). *Bikol grammar notes*. Honolulu: University of Hawaii Press.
- Mintz M W (1973). Case and semantic affixes of Bikol verbs. Ph.D. diss.: University of Hawaii.
- Mintz M W & del Rosario Britanico J (1985). *Bikol-English dictionary (with English-Bikol index)*. Manila: New Day Publishers.
- Mintz M W (1991). 'Anger and verse: two vocabulary subsets in Bikol.' In Harlow R (ed.) *VICAL2: Papers from the Fifth International Conference on Austronesian Linguistics, Western Austronesian and Contact Languages, I and II*. Auckland, New Zealand. Auckland: Linguistic Society of New Zealand.
- Owen N G (1999). *The Bikol blend: Bikolanos and their history*. Quezon City: New Day Publishers.
- Perez M R (1974). A critical study of Handiong, a Bikol epic. M.A. thesis: University of Santo Tomas, Manila.
- Realubit M L (1983). *Bikols of the Philippines*. Naga City: AMS Press.
- Realubit M L (1999). *Bikol literary history*. Naga City.
- Realubit M L & Erestain T E (1998). *Haliya: anthology of Bikol poets and poems*. Manila: National Commission for Culture and the Arts.
- San Agustin A de (1879). *Arte de la lengua Bicol para la enseñanza de este idioma*. Crespo M M (ed.) Manila: Tipographia de Ramírez y Giraudier, á cargo de C. Miralles.
- Santos P V (2003). *Hagkus: twentieth-century Bikol women writers*. Manila: De La Salle University Press.
- Stevens A M (1969). *Accent in Bikol*. Manila: Ateneo de Manila University.
- Yamada Y (1972). 'A Buhi Bikol vocabulary.' *Bulletin of the Faculty of Education* 24(1). Tokyo: Kochi University.

Bislama

C Hyslop, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

Bislama, an English-lexifier pidgin-creole, is the national language of Vanuatu, a republic in the south-west Pacific within the region of Melanesia. Along with English and French, it is also one of the official languages of the country. As the national language, it is spoken by the majority of the population as either a first or second language. There are as many as 100 distinct languages spoken in Vanuatu (81 actively spoken languages according to Lynch and Crowley, 2001) for a population of only 186 678 (1999 census), and as a result Bislama is vital as a lingua franca between speakers of different language groups. In urban areas and even in some rural areas, it is fast becoming the main language used in daily life. According to the 1999 census, in urban areas, where there is a great deal of intermarriage, Bislama is the main language used at home in 58% of households; in rural areas, this figure is considerably lower, at 13.3%. However, even in the most remote areas of the country only a minority of elderly people are not fluent in Bislama. Currently, English and French are the principal languages of education in Vanuatu and Bislama is generally banned in schools. However, Bislama is used for many other government and community services. For example, the majority

of radio broadcasts are in Bislama, although only some of the content of newspapers is published in Bislama. Parliamentary debates are conducted in the language, as are local island court cases.

Bislama is a dialect of Melanesian Pidgin, mutually intelligible with Solomons Pijin (Pijin), spoken in Solomon Islands, and Tok Pisin, spoken in Papua New Guinea. Thus, the language is not just an important lingua franca of Vanuatu, but also a common regional language that allows for communication among most peoples of Melanesia. Only in New Caledonia is Melanesian Pidgin not spoken.

The formation and development of Bislama, and of Melanesian Pidgin generally, took place within Vanuatu and other regions of Melanesia and also in Australia and other countries of the Pacific. A pidgin first started to emerge in Vanuatu (known as the New Hebrides at the time) in the mid-1800s as a result of the sandalwood and sea slug trade. Further development took place in the second half of the 19th century, with increasing numbers of Ni-Vanuatu being recruited to work on plantations both inside Vanuatu and in other areas of the Pacific, particularly in the sugarcane plantations of Queensland and Fiji (Crowley, 1990a). During the early decades of the 20th century, the language stabilized, such that its structure today is very close to what it was then. The status of and need for Bislama as a lingua franca within the country increased in the period leading up to independence in 1980, to the extent that today it has become the unifying language of the nation.

The majority of the Bislama lexicon, approximately 84–90%, is derived from English, reflecting its history of development alongside English-speaking traders, plantation owners, and colonists. Only approximately 3.75% of the vocabulary originates from the vernacular languages and 6–12% derives from French (Crowley, 2004). Of those words that derive from local languages, the majority describe cultural artifacts and concepts and endemic floral and faunal species that have no common names in English, such as *nasara* ‘ceremonial ground,’ *navele* ‘*Barringtonia edulis*,’ and *nambilak* ‘buff-banded rail.’ Note that many of these words start with *na-*, the form of an article or noun marker in many Vanuatu languages.

Although the majority of the lexicon is derived from English, the grammar of Bislama is greatly influenced by the vernacular languages. For example, in the pronominal system there is an inclusive-exclusive distinction in the first person, *yumi* ‘we (inclusive)’ is distinguished from *mifala* ‘we (exclusive).’ Dual and trial number is also distinguished from the plural, as *yutufala* ‘you (two),’ *yutrifala* ‘you (three),’ and *yufala* ‘you (pl.).’ Another feature that Bislama inherits from the substratum languages is reduplication. Reduplication is a productive process for both verbs and adjectives, but it is rarer for nouns. In verbs, reduplication can mark an action as being continuous, habitual, reciprocal, or random. It can mark intensity in both verbs and adjectives, and it also marks plurality in adjectives.

Like English and many Vanuatu languages, Bislama is characterized by AVO/SV word order, and this is the only means of recognizing the subject and object of the clause. Peripheral arguments are marked by prepositions. The preposition *long* has a wide general use; it marks the locative, allative, ablative, and dative. It can also mark the object of comparison in a comparative construction, the instrumental, and a number of other less easily defined functions. The preposition *blong* also has a

number of functions, marking the possessor in a possessive construction, a part-whole relationship, and a purposive role. Prepositions marking other semantic roles are *wetem* ‘with’ (instrumental and comitative), *from* ‘for, because of’ (reason), and *olsem* ‘like’ (similitive).

As is true of most pidgin languages, there is little marking of tense, aspect, and mood. The preverbal markers *bin* and *bae* mark the past and future tense, respectively. However, it is possible for an unmarked verb, preceded only by its subject, to indicate either past, present, or future tense, depending on the context. A number of auxiliaries also occur, with aspectual or modal functions, such as *stap*, marking a continuous or habitual action; *mas* ‘must’; *save* ‘be able’; and *wantem* ‘want.’ Verb serialization is a productive process in Bislama, encoding various meanings and functions such as a cause-effect relationship; a causative; or direction, position, or manner of action.

Bibliography

- Crowley T (1990a). *Beach-la-mar to Bislama: the emergence of a national language in Vanuatu*. Oxford Studies in Language Contact. Oxford: Clarendon Press.
- Crowley T (1990b). *An illustrated Bislama-English and English-Bislama dictionary*. Port Vila: University of the South Pacific, Pacific Languages Unit.
- Crowley T (2004). *Bislama reference grammar*. Honolulu: University of Hawai’i Press.
- Lynch J & Crowley T (2001). *Languages of Vanuatu: a new survey and bibliography*. Canberra, Australia: Pacific Linguistics.
- Tryon D T (1987). *Bislama: an introduction to the national language of Vanuatu*. Canberra, Australia: Pacific Linguistics.
- Tryon D T & Charpentier J-M (2004). *Pacific pidgins and creoles: origins, growth and development*. Trends in linguistics studies and monographs 132. Berlin: Mouton de Gruyter.

Brahui

P S Subrahmanyam, Annamalai University, Bangalore, India

© 2006 Elsevier Ltd. All rights reserved.

The word ‘Brahui’ designates both a language and its speakers. Brahui is the conventional spelling for the phonetically more correct Brāhōī/Brāhūī. The language is a member of the Dravidian family; more

specifically, it belongs to the North Dravidian subgroup, of which the other two members are Kurux and Malto. The Brahuīs live mainly in the Baluchistan and Sind provinces of Pakistan, but some are found also in Afghanistan (Šōrāwāk desert) and Iran (Sistan area). It is estimated that there are about 700 000 Brahui tribesmen, of whom only about 300 000 speak the language. Even those who speak Brahui are bilinguals in either Balochi or Siraki. There are

two views current among the scholars to explain the location of Brahui, which is far away from the main Dravidian area. Whereas one view maintains that the Brahuīs lived where they are now located from the earliest times, the other holds that they migrated to the current locations from that part of the main area that is occupied by the speakers of Kuṛux and Malto.

Phonology

The Brahui phonological system contains eight vowels and 28 consonants (see **Tables 1** and **2**). Proto-Dravidian short **e* and short **o* have been removed from the Brahui vowel system under the influence of Balochi; **e* developed into *ī/a* and **o* developed into *u/a/ō* (the exact conditionings are not known). The *ē* and *ō* have shorter (and somewhat lower) allophones before a consonant cluster.

The voiceless stops *p*, *t*, and *k* may optionally be accompanied by aspiration in all positions (*pōk*/*phōk*/*phōkh* ‘wasted’); however, aspirated stops in Indo-Aryan loans sometimes lose their aspiration in the south (*dhōbī/dōbī* ‘washerman’). The voiceless lateral *L* is the most characteristic sound of Brahui since it does not occur either in Proto-Dravidian (PDr) or in the neighboring languages of Brahui. It comes from two sources, PDr (alveolar) **l* and (retroflex) **ɭ*; both of these also show the reflex *l* in some words, the conditioning being unclear because of

the paucity of the data (*pāl* ‘milk’ < PDr **pāl*, *tēL* ‘scorpion’ < PDr **tēl*). The contrast between *L* and *l* is illustrated in *pāl* ‘milk’ and *pāl* ‘omen.’

One major dialectal division in Brahui involves the voiceless glottal fricative *h*; it appears in all positions in the northern dialects but is replaced in the south by the glottal stop in initial and intervocalic positions, and is lost before a consonant or in final position; the following examples illustrate the variation in the northern and southern dialects, respectively: *hust*, *ʔust* ‘heart’; *sahi affaṭ*, *saʔi affaṭ* ‘I don’t know’; *šahd*, *šad* ‘honey’; and *pōh*, *pō* ‘intelligence.’

Syntax

Word Classes

The following word classes may be recognized for Brahui: nouns (including pronouns and numerals), verbs, adjectives, adverbs (including expressives), particles, and interjections. An adjective normally occurs before the noun it qualifies but may be shifted to the postnominal position for the sake of emphasis:

jwān-ō hullī-as
good-INDEF horse-INDEF
‘good horse’

hullī-as jwān-ō
horse-INDEF good-INDEF
‘good horse’

Nouns and adjectives characteristically distinguish between definite and indefinite forms. The basic forms are definite and the corresponding indefinite ones are derived by adding -ō to the adjective base and -as to the nominal base, as illustrated in the preceding examples. A definite adjective that is monosyllabic is often strengthened by the addition of -ā/-angā:

sun-angā šahr
deserted village
‘deserted village’

Table 1 Vowels of Brahui

| | Front | | Central | | Back | |
|------|-------|------|---------|------|-------|------|
| | Short | Long | Short | Long | Short | Long |
| High | i | ī | | | u | ū |
| Mid | | ē | | | | ō |
| Low | | | a | ā | | |

Table 2 Consonants of Brahui^a

| | Labial | | Dental | | Alveolar | | Retroflex | | Palatal | | Velar | | Glottal (VL) |
|-----------|--------|----|--------|----|----------|----|-----------|----|---------|----|-------|----|--------------|
| | VL | VD | VL | VD | VL | VD | VL | VD | VL | VD | VL | VD | |
| Stop | p | b | t | d | | | ʈ | ɖ | c | j | k | g | ʔ |
| Nasal | | m | | | | n | | ɳ | | | | | |
| Fricative | f | | | | | | | | | | x | G | h |
| Sibilant | | | | | s | z | | | š | ž | | | |
| Lateral | | | | | L | l | | | | | | | |
| Trill | | | | | | r | | | | | | | |
| Flap | | | | | | ɾ | | | | | | | |
| Semivowel | | w | | | | | | | | y | | | |

^aAbbreviations: VD, voiced; VL, voiceless.

An indefinite adjective can function also as a noun:

ball-ō
big-INDEF
'big (one)'

An adverb occurs before the verb. Adverbs may be divided into those of (1) time (e.g., *dāsā* 'now,' *darō* 'yesterday,' *aynō* 'today,' *pagga* 'tomorrow'), (2) place (e.g., *monaṭī* 'forward'), and (3) manner (e.g., *dawn* 'thus'). For particles, the enclitic pronouns are very commonly used in Brahui. Whereas those for the third person are used in dialects throughout the Brahui area, those for the first and the second persons are more common in the Jahlawān dialect. They are suffixed to nouns or verbs. When added to a noun, they carry the sense of a pronoun in the genitive case; when added to a verb, they signal the direct or indirect object. The forms are: 1SG + *ka* 'my', 2SG + *nē* 'your,' 3SG + *ta* 'his/her/its', 3PL + *tā* 'their' (there are no plurals in the first and second persons):

maL-ē + ka
son-ACC/DAT + 1ENCL
'my son (accus.)/to my son'

xalkus + ka.
strike-PAST-2SG + 1ENCL
'You struck me.'

Word Order

The favored word order in Brahui is subject-object-verb:

I dā kārēmē kar-ōī uṭ
I this work do-NOM be.1SG
'I must do this work.'

Sentences Without the Copular Verb

Like most of the other Dravidian languages (especially the southern ones), Brahui contains sentences without the copula in certain contexts:

numā šahr-aṭī aṭ urā/ō?
your village-LOC how many house
'How many houses are there in your village?'

Gender and Number

Brahui, like Toda of South Dravidian, has no gender distinction, but number (singular versus plural) is distinguished (see later, Plural Suffixes). The original neuter forms (both singular and plural) of the third person are retained to refer to all categories: *ō(d)* 'he/she/it' (cf. Ta(mil). *atu* 'it', Te(legu). *adi* 'she, it') and *ōfk* 'they' (cf. Ta. *av(ay)*, Te. *avi* 'they (NEUT)').

Agreement

A finite verb shows agreement with the subject pronoun for person and number (see Table 3).

Noun Morphology

A nominal base is followed by the plural suffix when plurality has to be expressed and then by a case suffix; a postposition is normally attached to the genitive form of a noun.

Plural Suffix

The plural suffix is -k (variant -āk) in the nominative but -tē- before a nonnominative case suffix (see Table 4); as in the South Dravidian languages, use of the plural suffix is optional when plurality is understood from the context:

irā mār/mā-k (<*mār-k)
two son/son-PL
'two sons'

Case Suffixes and Postpositions

The nominative is unmarked; locative I means 'in' and locative II means 'on, by' (Table 4 shows all of

Table 3 Finite tenses of *tix-* 'to put'

| <i>Tense</i> | <i>Singular</i> | <i>Plural</i> |
|--------------------|------------------------------------|----------------------|
| Past | | |
| 1. | <i>tix-ā + t</i> 'I put' | <i>tix-ā + n</i> |
| 2. | <i>tix-ā + s</i> | <i>tix-ā + re</i> |
| 3. | <i>tix-ā</i> | <i>tix-ā + r</i> |
| Imperfect | | |
| 1. | <i>tix-ā + t-a</i> 'I was putting' | <i>tix-ā + n-a</i> |
| 2. | <i>tix-ā + s-a</i> | <i>tix-ā + re</i> |
| 3. | <i>tix-āk-a</i> | <i>tix-ā + r-a</i> |
| Pluperfect | | |
| 1. | <i>tix-ā + sut</i> 'I had put' | <i>tix-ā + sun</i> |
| 2. | <i>tix-ā + sus</i> | <i>tix-ā + sure</i> |
| 3. | <i>tix-ā + sas</i> | <i>tix-ā + sur</i> |
| Perfect | | |
| 1. | <i>tix-ā-n + ut</i> 'I have put' | <i>tix-ā-n + un</i> |
| 2. | <i>tix-ā-n + us</i> | <i>tix-ā-n + ure</i> |
| 3. | <i>tix-ā-n + e</i> | <i>tix-ā-n + a</i> |
| Present indefinite | | |
| 1. | <i>tix-i-v</i> 'I may put' | <i>tix-i-n</i> |
| 2. | <i>tix-i-s</i> | <i>tix-i-re</i> |
| 3. | <i>tix-e</i> | <i>tix-i-r</i> |
| Future | | |
| 1. | <i>tix-o-t</i> 'I will put' | <i>tix-o-n</i> |
| 2. | <i>tix-o-s</i> | <i>tix-o-re</i> |
| 3. | <i>tix-o-e</i> | <i>tix-o-r</i> |
| Nonpast negative | | |
| 1. | <i>tix-pa-r</i> 'I will not put' | <i>tix-pa-n</i> |
| 2. | <i>tix-p-ēs</i> | <i>tix-p-ēre</i> |
| 3. | <i>tix-p</i> | <i>tix-pa-s</i> |

Table 4 Case forms of *xal* 'stone'

| Case | Singular | Plural |
|-------------------|------------------|--------------------|
| Nominative | <i>xal</i> | <i>xal-k</i> |
| Accusative-dative | <i>xal-ē</i> | <i>xal-tē</i> |
| Instrumental | <i>xal-aṭ</i> | <i>xal-t-aṭ</i> |
| Comitative | <i>xal-tō</i> | <i>xal-tē-tō</i> |
| Ablative | <i>xal-ān</i> | <i>xal-tē-ān</i> |
| Genitive | <i>xal-nā</i> | <i>xal-tā</i> |
| Locative I | <i>xal-(a)ṭī</i> | <i>xal-tē-ṭī</i> |
| Locative II | <i>xal-ā(ī)</i> | <i>xal-tē-ā(ī)</i> |

the case forms of *xal* 'stone'). The following example shows postpositions:

ka-nā nēmaGāi
my towards
'towards me'

There are also a few prepositions, such as *bē(d)* 'without,' of Perso-Arabic origin that have entered Brahui through Balochi.

Pronouns

All of the pronouns are of Dravidian origin; however, Brahui developed postclitic forms of personal and demonstrative pronouns under the influence of Balochi (see preceding discussion, Word Classes). The first-person personal pronouns are *ī* 'I' and *nan* 'we'; the second-person personal pronouns are *nī* 'you(singular)' and *num* 'you (plural)'. There is only the singular reflexive pronoun, *tēn* 'self'. The interrogative pronouns are *dēr* 'who?' and *ant* 'what?'. The third-person forms show a threefold deictic distinction: proximal *dā(d)* '(one) who is here' (plural *dāfk*), medial *ē(d)* '(one) who is at some distance' (plural *ēfk*), and distal *ō(d)* '(one) who is far off' (plural *ōfk*).

Numerals

Only the cardinal numbers for one, two, and three are of Dravidian origin (the forms without the final *ṭ* of these function as adjectives); all others are borrowed from Balochi. The number '1' is *asi(t)*, '2' is *ira(t)*, and '3' is *musi(t)*.

Verb Morphology

Verb Bases

A verb base in Brahui may be simple or complex. The complex base is formed from the simple one by the addition of the transitive-causative suffix *-if* (conditioned variant: *-f*). This suffix converts an intransitive into a transitive and an underived transitive into the corresponding causative; it is, therefore, possible to use the suffix twice in a sequence,

e.g., *bin-* 'to hear,' *bin-if-* 'to cause to hear,' *ka?* 'to die,' *kas-f-* 'to kill,' and *kas-f-if-* 'to cause (someone) to kill.'

Finite Verbs

There are four kinds of past tense (past, imperfect, pluperfect, and perfect), each with different shades of meaning, and all of them are periphrastic constructions involving the 'be' verb. The past stem, which is the basis for all of these, is formed by adding to the base *-b-* (conditioned variants: *-ē-*, *-k-*, *-g-*, *-is-*, *-s-*, *-ss-*). The following formulas give the structures of these tenses:

1. Past: past stem + present of *ann-* 'to be.'
2. Imperfect: past + *a*.
3. Pluperfect: past stem + past of *ann-* 'to be.'
4. Perfect: past stem + (*u*)*n* + present of *ann-* 'to be.'

The present indefinite, the future, and the nonpast negative are morphological constructions with the following structures (these and the previously mentioned tenses are illustrated in Table 3 with the verb base *tix-* 'to put'):

1. Present indefinite: verb base + *i* + personal suffix.
2. Future: verb base + *o* + personal suffix.
3. Nonpast negative: verb base + *pa* + personal suffix.

There are some other syntactic constructions involving *ann-* 'to be' that need not be mentioned here. One noteworthy feature of Brahui is the strategy of suffixing *-a* to form one type of finite verb from another. The imperfect present-future and the negative present-future are thus formed from the past present-indefinite and the nonpast negative, respectively.

The imperative suffixes are 2SG *-ø*, 2PL *-bo* (conditioned variant: *-ibo*):

tix
put-2SG
'Put!'

tix-bo
put-2PL
'Put (plural)!'

The corresponding negative imperative has the negative suffix *-pa-* (conditioned variant: *-fa-*) between the base and the imperative suffix:

tix-pa
put-NEG-2SG
'Don't put (singular)!'

tix-pa-bo
put-NEG-2PL
'Don't put (plural)!'

Nonfinite Verbs

The present adverb has the suffix *-isa*:

bis-isa
bake-PRES ADV
'baking'

The present adjective has the suffix *-ok*:

bin-ok
hear-PRES ADJ
'that hear(s)'

The infinitive-cum-action noun is formed by adding *-ing* (conditioned variant: *-ēng*) to the verb base:

bin-ing
hear-INF/VN
'to hear, hearing'

Bibliography

- Bausani A (1969). 'La letteratura Brahui.' In Botto O (ed.) *Storia delle letterature d'Oriente II*. Rome. 649–657.
- Brahui A R (1983). 'History, background, objectives and achievements of the Brahui Academy, Quetta, Pakistan.' In Rossi A & Tosi M (eds.) *Newsletter of Baluchistan studies I*. Naples.
- Bray D (1909). *The Brahui language I*. Calcutta [Reprinted in 1972 in Quetta].
- Bray D (1913). *Life-history of a Brahui*. London.
- Bray D (1934). *The Brahui language II: The Brahui problem*. Delhi [Reprinted in 1978 in Quetta].
- Bray D (1939). 'Brahui tales.' *Acta Orientalia* 17, 65–98.
- DeArmond R (1975). 'Some rules of Brahui conjugation.' In Schiffman H & Eastman C (eds.) *Dravidian phonological systems*. Seattle: University of Washington. 242–298.
- Elfenbein J (1982). 'Notes on the Balochi-Brahui commensality.' *Transactions of the Philological Society*. 77–98.
- Elfenbein J (1983). 'The Brahui problem again.' *Indo-Iranian Journal* 25, 103–132, 191–209.
- Elfenbein J (1987). 'A periplus of the Brahui problem.' *Studia Iranica* 16, 215–233.
- Elfenbein J (1997). 'Brahui phonology.' In Kaye A (ed.) *Phonologies of Asia and Africa*. Winona Lake, IN: Eisenbrauns. 797–811.
- Elfenbein J (1998). 'Brahui.' In Steever S B (ed.) *The Dravidian languages*. London and New York: Routledge. 388–414.
- Emeneau M B (1937). 'Phonetic observations on the Brahui language.' *Bulletin of the School of Oriental and African Studies* 8(4), 981–983.
- Emeneau M B (1962a). *Brahui and Dravidian comparative grammar*. Berkeley: University of California Publications in Linguistics.
- Emeneau M B (1962b). 'Bilingualism and structural borrowing.' *Proceedings of the American Philosophical Society* 106, 430–442.
- Emeneau M B (1991). 'Brahui personal pronouns, first singular and reflexive.' In Bai B L & Reddy B R (eds.) *Studies in Dravidian and general linguistics (a Festschrift for Bh. Krishnamurti)*. Hyderabad: Osmania University. 1–12.
- Grierson G (1906). *Linguistic survey of India, vol. 4: the Munda and Dravidian languages*. Calcutta.
- Grierson G (1921). *Linguistic survey of India, vol. 10: Eranian family*. Calcutta.
- Mayer T J L (1906–1907). *A Brahui reading book* (vols. I–III). Ludhiana [Reprinted in one volume in 1983 by the Brahui Academy, Quetta].
- Tate G P (1909). *The frontiers of Baluchistan*. London.

Breton

J Le Dû, Université de Bretagne Occidentale,
Brest, France

© 2006 Elsevier Ltd. All rights reserved.

Breton (*brezoneg*, *brezhoneg*) belongs to the Brythonic branch of the Celtic languages. It is spoken in Lower Brittany, and its linguistic border is the westernmost limit of the withdrawal of Celtic before Roman expansion.

Breton has long been considered the continuation of Gaulish. Linguistic studies in the 19th century smothered all purported genetic connection between Breton and French and also any close relationship to Gaulish. Some historians argued that Breton had been imported whole by immigrants from Britain into

a thoroughly romanized Armorica. Modern Celtic studies confirmed the view that Breton was a late offshoot of British Celtic. We now know that emigration from Britain began before the Saxon invasions, so that most scholars acknowledge that Breton is rooted in Armorican Gaulish, absorbing different varieties of British Celtic.

A traditional view of the language purports the existence of a unified old Breton, supposed to have split into four dialects, named after the dioceses as they existed before the 1789 French Revolution: Léonais for the diocese of Léon, Trégorrois for Tréguier, Cornouaillais for Cornouaille, and Vannetais for Vannes. There are, in fact, two major dialect groups: (1) KLT – Cornouaille (Kerne), Léon, Trégor and (2) Vannetais, the western border of which is the river

Ellé. Falc'hun (1962, 1981) has reported the existence of an intermediate dialect centered on Carhaix, the meeting point of all the major roads, and constituting a bridge between remote linguistic forms, like the reflexes of the dental spirants from old Celtic **tt* and **d*. Léon *deiz* 'day' and *dervez* 'duration of a day' (Welsh *dydd* and *dyddwaith*) are far removed from vannetais *de* and *deñeh*. The central forms are *de* and *devez*, dropping *z* from **d* as in vannetais, but keeping *z* from **tt* as in Léon. The primitive twofold partition could reflect the difference between Osismii and Venetes Gaulish, the latter keeping closer to Armorican.

An intensity stress generally falls on the penultimate in the northwest, whereas in the Southeast a pitch stress affects the last syllable, not unlike French.

Voiceless consonants and /m/ are fortes, voiced spirants are lenes, and voiced stops and /l/, /n/ and /r/ can be either. Vowels are short before fortes and long before lenes when stressed. One can thus oppose *ar zal* 'the room' (long [a:], weak [l]) and *zall* 'salted' (short [a], strong [l]). There can be up to eight phonemic nasal vowels, which are not borrowings from French, but archaic features, as in *hañv* 'summer'.

Primitive consonants were weakened, especially between vowels. These changes survived the loss of final syllables, turning a simple phonetic mechanism into a grammatical device called 'lenition,' so that the initial consonants of feminine words are lenited after the article – originally ending in a vowel – and also the following adjective: *mamm* 'mother,' *mad* 'good,' *ar vamm vad* 'the good mother.' The geminate voiceless fortes became voiceless spirants, giving rise to the spirant mutation: *penn* 'head,' *he fenn* 'her head.' Another sandhi phenomenon caused the so-called protective mutation: a final *-h* in *hoh* 'your' devoiced a following voiced initial consonant, as in *bugel* 'child,' *ho(h) pugel* 'your child.'

Final consonants are devoiced before pause. *Ma zad* 'my father' keeps a long [a:], but a devoiced *-d* when final, the voice being restored when the utterance is followed by a vowel as in *ma zad eo* '(he) is my father'. All voiceless consonants are voiced before a vowel or *l*, *m*, *n*, and *r*. Native Breton speakers are readily recognizable in French when they pronounce *toud' la z'maine* for *toute la semaine* 'during the whole week.'

English and Breton grammars show striking similarities; for example, both use a compulsory periphrastic progressive in opposition to a simple present: *Ma breur ne gan ket* 'my brother does not sing' vs. *ma breur n'ema ket o kana* 'my brother is not singing.'

The lexis is basically Celtic (*dorn* 'fist, hand', Welsh *dwrn*, Gaelic *dorn*; *den* 'person', Welsh *dyn*, Gaelic *duine*). About 500 common words are Latin borrowings (*taol* < *tabula* 'table,' *spered* < *spiritus* 'mind,'

kistin < *castanea* 'chestnut'). For centuries, a flow of romance and French words has enriched the language, very much like in English. Some words have been kept in both languages while disappearing from French; for example, *skourje* 'whip' from *escourgée*, *écourgée*, English *scourge*. The most important borrowings are the numerous affixes, taken both from Latin (*-adur* < *-atura* as in *skub-adur* 'sweepings') and French (*lenn-abl* 'read-able').

Polls carried out in 1991 and 1997 show that from 1950 to 1990, the percentage of Breton speakers has decreased from about roughly 70 to 20% of the population. In 2004, it is estimated that about 250 000 persons are able to speak the language, and most of them are over 60 years old. French has become dominant because of the unprecedented social and agricultural revolution occurring in Brittany.

Before 1941, there existed two written forms, called at the time 'breton vannetais' and 'bas-breton,' which had been developed in the two Jesuit colleges of Quimper and Vannes in the 17th century; each form had its own grammars, dictionaries, and literature. In 1941 the *peurunvan* 'totally unified' orthography was established. 'Cat,' *kaz* in KLT and *kah* in vannetais, would be spelled *kazh*. A new spelling called 'orthographe universitaire,' which was closer to the spoken language, was created in 1954. Finally, a third orthography, *etrerannyezhel* 'interdialectal,' was created in the 1970s to take into account all regional differences.

Both the French State and the Breton Regional Assembly have encouraged publishing in the Breton language in the last 30 years, and Breton is partially used on local state-owned (France-Bleu Breiz Izel) and private radio (like *Radio Kerne*) and television stations (France 3).

Degrees in Breton, at all levels, are delivered in Rennes and Brest. Breton language teachers have been recruited since 1982 to teach in the secondary schools. Breton is taught to about 5000 children at the primary level in a few bilingual classes in public and Catholic schools, and the private *Diwan* schools teach mostly through Breton. However, less than 1% of Breton children benefit from this bilingual education.

Bibliography

- Balcou J & Le Gallo Y (eds.) (1987). *Histoire littéraire et culturelle de la Bretagne*. Paris-Genève: Champion-Slatkine.
- Broudic F (1995). *La pratique du breton de l'ancien régime à nos jours*. Rennes: PUR.
- Falc'hun F (1962). 'Le Breton, forme moderne du gaulois, Rennes.' *Annales de Bretagne* 64(4).

- Falc'hun F (1981). *Perspectives nouvelles sur l'histoire de la langue bretonne*. Paris: UGE.
- Fleuriot L (1980). *Les origines de la Bretagne*. Paris: Payot.
- Guimard J-Y (1987). *Le Bretonisme: les historiens bretons au XIX^e siècle*. Société d'Histoire et d'archéologie de Bretagne.
- Hélias P-J (1979). *The horse of pride – life in a Breton village*. London/New Haven: Yale University Press.
- Hemon R (1975). *A historical morphology and syntax of Breton*. Dublin: The Dublin Institute for Advanced Studies.
- Hersart de La Villemarqué T (1867). *Barzaz Breiz – chants populaires de la Bretagne*. Paris: Perrin.
- Humphreys L I H (1995). *Phonologie et morphosyntaxe du parler breton de Bothoa en Saint-Nicolas-du-Pelem*. Brest: Emgleo Breiz-Brud Nevez.
- Jackson K (1967). *A historical phonology of Breton*. Dublin: Dublin Institute for Advanced Studies.
- Le Berre Y & Le Dû J (1997). 'Nommer le breton.' In Tabouret-Keller (ed.) *Le nom des langues I: les enjeux de la, nomination des langues*. Louvain-La-Neuve: BCILL, Peeters.
- Le Berre Y & Le Dû J (1999). 'Le qui pro quo des langues régionales: sauver la langue ou éduquer l'enfant?' In Clairis C, Costaouec D & Coyos J-B (eds.) *Langues et cultures régionales de France – Etat des lieux, enseignement, politiques*. Paris: L'Harmattan.
- Le Dû J (2001). *Nouvel atlas linguistique de la Basse-Bretagne*. Brest: CRBC.
- Le Roux P (1924–1963). *Atlas linguistique de la Basse-Bretagne*. Rennes and Paris.
- Loth J (1883). *L'immigration bretonne en Armorique du V^eme au VII^eme siècles de notre ère*. Paris.
- McKenna M (1988). *A handbook of modern spoken Breton*. Tübingen: Niemeyer.
- Piette J F R (1973). *French loanwords in Middle Breton*. Cardiff: University of Wales Press.
- Ploneis J-M (1983). *Au carrefour des dialectes breton: le parler de Berrien*. Paris: SELAF.
- Sommerfelt A (1978). *Le breton parlé à Saint-Pol-de-Léon: phonétique et morphologie* (2^eme éd.). Oslo: Universitet-forlaget.
- Tanguy B (1977). *Aux origines du nationalisme breton-le renouveau des études bretonnes au XIX^eme siècle*. Paris: UGE.
- Ternes E (1970). *Grammaire structurale du breton de l'île de Groix*. Heidelberg: Carl Winter.

Bulgarian

J Miller, University of Auckland, Auckland, New Zealand

© 2006 Elsevier Ltd. All rights reserved.

Bulgarian is a South Slavic language, along with Slovene (Slovenian), Macedonian, and the Serb-Croatian linguistic complex. Geographically Bulgarian is also a Balkan language and shares a number of phonetic, grammatical, and lexical features with Rumanian (Romanian), Greek, and Albanian. For instance, Rumanian and Albanian have schwa in stressed syllables and so does Bulgarian, the only Slav language with this property.

Bulgarian has two sets of dialects, Eastern and Western (further subdivisions are recognized). A major difference is in the reflexes of the Common Slavic *jat* vowel, roughly equivalent to 'ye' as in English *yet*. In the North Eastern dialects the *jat* vowel became 'ja' in a stressed syllable and followed by a syllable with a back vowel. Elsewhere it became 'e.' Standard Bulgarian, based on the North Eastern dialects, has the 'ja' – 'e' alternation, in, e.g., adjectives: *bjalo* 'white' (neuter singular) versus *beli* (plural).

The Common Slavic 'l' and 'r' plus *jer* (extra-short vowel) and syllabic 'l' and 'r' became 'ŭr' and 'ŭl' in polysyllabic words before two consonants and 'rŭ' and 'lŭ' elsewhere: *skŭrben* 'sorrowful'; 'prŭv' (first-person

masculine) versus 'pŭrva' (first-person feminine). Consonants are palatalized or non-palatalized, as in other Slav languages.

Bulgarian has lost the Slavic case-suffixes but has developed definite articles, attached to the first word in noun phrases: Bulgarian *knigata* 'the book,' *kniga* 'a book,' *novata kniga* 'the new book,' *nova kniga* 'a new book.' In written Bulgarian masculine nouns take different subject and oblique forms of the article: (*j*)*at* and (*j*)*a*. In spoken Bulgarian (*j*)*at* is typically not used.

Bulgarian has preserved the Indo-European tense-aspect system of imperfect and aorist alongside the newer perfective-imperfective system. Typically, imperfect suffixes are added to imperfective stems and aorist suffixes to perfective stems. Bulgarian does offer examples of perfective stems with imperfect suffixes in subordinate clauses introduced by, e.g., *shtom* 'as soon as' and in main clauses; they express a completed action that is repeated. The following example (1) is from Feuillet (1995: 36).

- (1) Večer sedneshe na chardaka
Evening sit-down-3SG on verandah-DO
'In the evening he would sit down on the verandah'

Sedn is perfective and *-eshe* is imperfect.

(him-accusative), and so on. Consider the question–answer pair in (9).

- (9) Chete li ja Dimo novata kniga?
 Read Q it Dimo new-the book?
 ‘Did Dimo read the new book?’
 Dimo ja chete novata kniga
 Dimo it read new-the book
 ‘Dimo read the new book’

(9) is neutral; it asks simply if this event took place, not whether it was Dimo doing it or someone else, or if it was the new book that was read or something else. The order *novata kniga ja chete Dimo* highlights *ja chete Dimo*; the pronoun *ja* signals that *novata kniga* is the direct object of *chete*. The order *novata kniga, Dimo ja chete*, with focal stress on *Dimo*,

puts contrastive highlighting on *Dimo*: ‘As for the book, it was Dimo who read it and not anyone else.’

Bibliography

- Aronson H (1968). *Bulgarian inflectional morphophonology*. The Hague: Mouton de Gruyter.
 De Bray R G A (1980). *Guide to the Slavonic languages* (3rd edn.). Columbus, OH: Slavica.
 Feuillet J (1995). *Bulgare*. Munich: Lincom Europa.
 Holman M & Kovacheva M (1993). *Teach yourself Bulgarian*. London: Hodder and Stoughton.
 Hubenova M, Dzhumadanova A & Marinova M (1983). *A course in modern Bulgarian, parts 1 and 2*. Columbus, OH: Slavica.
 Rudin C (1985). *Aspects of Bulgarian syntax: complementizers and wh constructions*. Columbus, OH: Slavica.

Burmese

J Watkins, School of Oriental and African Studies,
 London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Burmese is the national language of Burma/Myanmar and is the mother tongue of the Burman (Bamar) ethnic majority, who make up approximately two-thirds of Burma’s population of slightly over 50 million. The rest of the country’s indigenous population is diverse, speaking between 60 and 100 other languages among them, depending on the criteria used to distinguish languages from one another. Most non-Burmans live in the areas near Burma’s borders with Thailand, Laos, China, India, and Bangladesh, although many live interspersed with Burmans and speak Burmese and other languages in addition to their native language. Burmese is little spoken outside Burma, but widely dispersed and fragmented communities of Burmese expatriates may be found in Asia and around the world.

Burmese belongs to the Tibeto-Burman language family, which comprises approximately 350 languages spoken across a vast territory stretching from the Himalayas to mainland Southeast Asia. Burmese has by far the largest number of speakers of any of the Tibeto-Burman languages, most of which have only a few thousand speakers and many of which may disappear during the 21st century.

Most of the other languages spoken in Burma also belong to the Tibeto-Burman language family. Some,

such as Arakanese (Rakhine), Intha, and Danu, are so similar to Burmese as to be considered by some to be dialects of Burmese rather than separate languages.

History and Script

The Burmese have been in the area of modern Burma/Myanmar from approximately 850 C.E. onward, founding their capital at Pagan (Bagan). Despite extensive contact over the following two centuries with the Pyu, the speakers of a now-dead Tibeto-Burman language that occupied the area, the first inscriptions in Burmese date from the 11th century, with no extant examples of Burmese writing before then. Burmese script is a close cousin of the Mon script, which was adapted from a southern Indian script, a descendant of the Brāhmī script that was the ancestor of many Indic scripts found in South and Southeast Asia. It is thought that the Burmese adapted the script from Mon after Mon scribes were brought to the city of Pagan after the Burmese king Anawratha, in 1057 C.E., defeated the Mon, although this theory has been disputed in recent research.

Aside from the rounding of the originally square characters into the distinctive round-shaped letters of Burmese today, the alphabet has remained largely unchanged to the present day. It is widely believed that the round shapes of Burmese letters evolved because texts were traditionally written on palm leaves, which would split easily if angled shapes were scratched on them. Whether or not this is true, Burmese writing retains its distinctive round shapes, and handwriting with consistent, even circles is praised.

The writing system evolved between the period of the early inscriptions and the 16th century C.E. when it assumed a form similar to its present-day state. The spoken language has changed considerably since that time, with the result that a faithful transliteration of written Burmese (such as the one approved by the American Library Association and the Library of Congress used here) gives little impression of the way letters or words are pronounced in the language today. Sound changes have applied to certain initial consonants. Final consonants have disappeared. A glottal stop is all that remains of final stop consonants, whereas the place contrasts of written final stops are realized as vowel changes in the syllable. Final nasal consonants have been replaced by a parallel series of nasalized vowels. In general, many combinations of symbols are pronounced differently from the sounds represented by the symbols individually.

The phonetic transcription used here is faithful to the principles of the IPA, although several others have been devised. A transliteration and transcription are compared in the following example.

Burmese script ရုပ်မြင်သံကြား
 Transliteration RUP'MRAN'SAMKRĀ
 Transcription jouʔ.mjɪn.θàn.dʒá
 Gloss picture.see.sound.hear
 Translation 'television' (more commonly တီဗွီ tí.vì
 'T.V.')

Burmese script is basically alphabetic. There are separate symbols to represent consonants (Table 1) and vowels (Table 2), but the symbols are organized

in syllabic clusters, which are written from left to right. Within each cluster, however, the symbols do not necessarily appear in left-to-right order. For example, to write the syllable တီ tí 'worm,' the vowel ိ-ì is placed on top of the consonant တ t, but to write တူ tí 'nephew,' the ိ-ù must hang below the initial တ t. Certain sounds in Burmese, namely affricates, voiceless sonorants, and initial consonant clusters, are written using medial forms of four consonants, shown in Table 3.

Burmese script has retained the features and symbols needed for writing the South Asian languages for which its parent scripts were originally designed, such as Pāli, the language of the Buddhist scriptures and the source of many loans in Burmese, which can easily be identified because of phonological features such as doubled consonants and retroflex consonants that do not occur in Burmese words. A Pāli phrase and its rendition in Burmese are shown next.

Burmese script
 Transliteration
 Transcription

ဗုဒ္ဓံသရနံဂစ္ဆာမိ
 BUDDHĀM SARANĀM GACCHĀMI
 bouʔdàn θaɹənàn gjiʔsʰamɪ
 'I go to the Buddha for refuge'

Phonetics and Phonology

Some of the sounds used in Burmese are considered unusual because they occur relatively rarely in the world's languages. These are the so-called voiceless nasals, which include the sound of air escaping through the nose. The Burmese word for ရင်းနှီးမြှုပ်နှံမှု jín.nù. mjouʔ.nàN.mū 'investment' contains examples

Table 1 Consonants of Burmese, transliterated and transcribed

| Burmese script | | | | | Tranliteration | | | | | Transcription | | | | |
|----------------|---|---|---|---|----------------|----|---|----|-------|---------------|-----------------|---|---|---|
| က | ခ | ဂ | ဃ | င | K | KH | G | GH | Ń | k | k ^h | g | g | ŋ |
| စ | ဆ | ဇ | ဈ | ည | C | CH | J | JH | Ñ /ÑÑ | tɕ | tɕ ^h | ɕ | ɕ | ɲ |
| ဋ | ဌ | ဍ | ဎ | ဏ | Ṭ | ṬH | Ḍ | ḌH | Ṇ | t | t ^h | d | d | n |
| တ | ထ | ဒ | ဓ | န | T | TH | D | DH | N | t | t ^h | d | d | n |
| ပ | ဖ | ဗ | ဘ | မ | P | PH | B | BH | M | p | p ^h | b | b | m |
| ယ | ရ | လ | ဝ | သ | Y | R | L | V | S | j | r | l | w | θ |
| ဟ | ဠ | အ | | | | H | Ḷ | - | | | h | l | ʔ | |

of two such sounds: /m̥/ and /m̥/. The consonants in Burmese are set out in **Table 4**.

For reasons of historical phonology, vowels in orthographically open syllables (**Table 5**), which are written with no final consonant letter, can be distinguished from those found in orthographically closed syllables (**Table 6**) namely those ending in a glottal stop or with a nasal vowel (transcribed here with /N/, which does not represent a final nasal consonant), both of which are written as final consonant letters in the writing system.

Like the majority of the languages spoken in mainland East and Southeast Asia, Burmese is a

tone language. The tonal contrasts involve not only the commonly observed differences in pitch and vowel length but also differences in phonation type – whether the voice is breathy or sharp in character. The presence or absence of a glottal stop at the end of the syllable may also be considered to be part of the tonal system. **Table 7** gives a basic description of the tonal contrasts on a syllable consisting of a bilabial nasal and an open vowel.

Burmese morphemes in phrases and compounds display varying degrees of phonological juncture, principally voicing assimilation and reduction of the first syllable, as shown in the following examples.

- Voicing assimilation on internal morpheme boundaries in compounds.

| | |
|----------------------|---|
| ပန်း + ခြံ > ပန်းခြံ | pán + tɕʰán > 'flower' + pándzán 'enclosure' > 'garden' |
| စား + ပွဲ > စားပွဲ | sá + pwé > 'eat' + 'event' > sábwé 'feast' |
| မြင် + တယ် > မြင်တယ် | mjìn + tè > 'see' + REALIS > mjìndè 'sees/saw' |

- Reduction of first element in compounds.

| | |
|-----------------------|---|
| ခါ + ပိုက် > ခါးပိုက် | kʰá + pai? > waist + carry > gəbai? 'pocket' |
| စား + ပွဲ > စားပွဲ | sá + pwé > eat + event > zəbwé 'table' |

Table 2 Burmese word-initial and word-internal vowel symbols

| Word-initial | | Word-internal | | Transliteration | |
|--------------|----|---------------|----|-----------------|---|
| အ | အာ | — | — | A | Ā |
| ဣ | ဩ | ◌◌ | ◌◌ | I | Ē |
| ဥ | ဦ | ◌◌ | ◌◌ | U | Ū |
| ဧ | ဩ | ◌◌ | ◌◌ | E | O |

Table 3 Medial forms of Burmese consonants

| Initial | Medial | Burmese | Transcription | Pronunciation | Gloss |
|---|--------|---------|---------------|---------------|-----------------|
| ယ | ၂ | များ | MYĀ" | mjá | ‘many’ |
| | | ကျား | KYĀ" | tɕá | ‘tiger’ |
| ရ | ၉ | မြန်မာ | MRAN‘MĀ" | mjànmà | ‘Burma/Myanmar’ |
| | | ခြောက် | KHROK‘ | tɕʰauʔ | ‘dry’ |
| ဝ | ◌◌ | လွယ် | LVAY‘ | lwè | ‘easy’ |
| | | နွား | NVĀ" | nwá | ‘cow’ |
| ဟ | ၂ | မှာ | MHĀ" | mà | ‘order’ |
| | | မှုံ | RVHĀM‘ | ʃuɴ | ‘mud’ |
| Words may be spelled with a maximum of three medial consonants: | | | | | |
| | | အမြွှာ | AMRVHĀ | ʔəmɰwà | ‘segment’ |

Table 4 The consonants of Burmese

| | <i>Bilabial</i> | <i>Dental</i> | <i>Alveolar</i> | <i>Postalveolar</i> | <i>Palatal</i> | <i>Velar</i> | <i>Glottal</i> |
|--------------------|--------------------|---------------|--------------------|----------------------|----------------|--------------------|----------------|
| <i>Stop</i> | p p ^h b | | t t ^h d | tɕ tɕ ^h ɖ | | k k ^h g | ʔ |
| <i>Nasal</i> | m m̥ | | n ɳ | | ɲ ɲ̥ | ŋ ɲ̥ | |
| <i>Fricative</i> | | θ ð | s s ^h z | ʃ | | | h |
| <i>Approximant</i> | w ʍ | | (ɹ) | | j | | |
| <i>Lateral</i> | | | l ɭ | | | | |

Table 5 Vowels of Burmese in orthographically open syllables

| Vowel | Lowe tone | | High tone | | Creaky tone | |
|-------|-----------|--|-----------|--|-------------|--|
| /i/ | | | | | | |
| /e/ | | | | | | |
| /ɛ/ | | | | | | |
| /a/ | | | | | | |
| /ɔ/ | | | | | | |
| /o/ | | | | | | |
| /u/ | | | | | | |

Morphology

Morphemes in Burmese are predominantly monosyllabic. With the exception of Indo-European loans, typically from Pali or English, compounding is the major source of polymorphemic words. In the television example above, four morphemes (N + V) (N + V) combine to form a noun.

Derivational morphology by prefixation is common, in particular noun-formation from verbs using the prefix ʔə-ʔə-.

| | | |
|-----------------------------|--|--------------------------|
| ပြိုင်ဆိုင် > အပြိုင်အဆိုင် | pjains ^h ain > ʔəpjain ʔas ^h ain | compete > competition |
| ရောင်း / ဝယ် > အရောင်းအဝယ် | jaun / wè > ʔəjaun ʔawè | sell / buy > trade |

The verbal complex, typically occurring at the end of a Burmese sentence, may comprise one or more head verbs in series followed by a string of auxiliary verbs, verbal particles, and markers.

| | | |
|-----------------------------|-----------|--------------------------|
| NP | NP | VP |
| ခေတ်မီဈေးတွေ | ဟိုတယ်တွေ | ဖြစ်ပေါ်လာစေပြန်ပါတယ်။ |
| k ^h i?mì. zé.dwè | hòtè.dwè | p ^h i?pò. là. |

modern.market.PL hotel.PL zè.bjàN.bà.dè
become emerge.begin.

CAUS.also.

POLITE.REALIS

‘... caused modern markets and hotels to begin to appear as well’

Burmese has a system of noun case markers, which in many contexts are not obligatorily present, and postpositions, as illustrated next.

| | | | |
|-----------|-------------|-------------|-----------|
| ဦးဘက | မန္တလေးကို | အမေနဲ့ | သွားတယ်။ |
| Yu.ba_ga | māndalé.gò | ʔəmè.nɛ | θwá.dè |
| U Ba.SUBJ | Mandalay.to | mother.with | go.REALIS |

Burmese, like other languages of the region, encodes power and solidarity in personal relationships using a rich system of pronouns and forms of address. Pronouns may be true pronouns, such as ငါ 1SING ‘I’ and နင် nɪ̃n 2SING ‘you’ (both familiar, not polite), or grammaticalized from other sources, such as ကျွန်တော် tɕə̀nɔ̃ 1SING (male, polite; literally ‘royal slave’). Other forms of address include titles, personal relationships, and names or a combination of all three, such as ဆရာမခေါ်ခင်ခင်ချော sʰəjamǎ. dɔ̌.kʰĩ nkʰĩntɕʰə̌ ‘Teacher (FEM) Aunt (= Mrs.) Khin Khin Chaw.’

Literacy and Literary Burmese

The literacy rate in Burma has often been said to be high compared to other countries in the region, but accurate data are extremely difficult to obtain. One recent source suggests that nearly 80% of Burmese

Table 6 Vowels of Burmese in orthographically closed syllables: killed tone or nasal vowel

| Vowel | Nasal vowel | | | | | | Killed tone | |
|-------|-------------|------|-----------|------|-------------|------|-------------|-----|
| | Low tone | | High tone | | Creaky tone | | | |
| /i/ | —ိ | ိၤ | —ိး | ိၤ | —ိ့ | ိၤ | —ိ် | iʔ |
| /e/ | —ေိ | ဲိၤ | —ေိး | ဲိၤ | —ေိ့ | ဲိၤ | —ေိ် | eiʔ |
| /ɛ/ | | | | | | | —က် | ɛʔ |
| /ai/ | —ါိ | ဲါိၤ | —ါိး | ဲါိၤ | —ါိ့ | ဲါိၤ | —ါိ် | aiʔ |
| /a/ | —ာ | ာၤ | —ား | ာၤ | —ာ့ | ာၤ | —ာ် | aʔ |
| /au/ | —ော | ဲောၤ | —ေား | ဲောၤ | —ော့ | ဲောၤ | —ော် | auʔ |
| /ou/ | —့ | ့ၤ | —့း | ့ၤ | —့့ | ့ၤ | —့် | ouʔ |
| /u/ | —ု | ုၤ | —ုး | ုၤ | —ု့ | ုၤ | —ု် | uʔ |

Table 7 Burmese tones

| Tone name | IPA | Burmese script | Description |
|-----------|-------|----------------|--------------------------------------|
| Low | [ma:] | မာ | long, low |
| High | [ma:] | မား | long, high; sometimes breathy |
| Creaky | [mǝ] | မ / မ့ | short, high, with creaky phonation |
| Killed | [maʔ] | မတ် | short, high, with final glottal stop |

Syllables with one of these tones may in some contexts become reduced to a short, unstressed schwa which is counted as a fifth tonal category in some analyses.

people over the age of 15 are literate, but other sources have put the figure much lower.

The Burmese language exists in a colloquial style used in spoken informal contexts and a literary style used in official formal settings. The main difference between the two is that they have separate sets of grammar words and some other vocabulary. A colloquial-style sentence is compared to its literary-style equivalent in the next example.

Spoken ဦးဘက မန္တလေးကို အမေနဲ့ လာတယ်။
 ʔú.ba_ga_mánda.lé.gò ʔə.mè.nɛ lá.dɛ̌
 Literary ဦးဘသည် မန္တလေးသို့ အမေနှင့် လာ၏။
 ʔú.ba_ði mánda.lé.ðò ʔə.mè.ɲin lá.ʔi
 U Ba.SUBJ Mandalay.to mother.with come.
 REALIS
 ‘U Ba came to Mandalay with his mother’

Given the large number of speakers of Burmese and the existence of a large diaspora community

scattered around the world, Burmese has an inevitable presence on the Web, although at the time of writing standardized encoding has yet to be widely adopted and so text is usually displayed on the Internet as graphics. For ease of use, computer users often render Burmese in romanized form in Internet chat rooms or e-mail.

Bibliography

- Allott A (1985). 'Language policy and language planning in Burma.' In Bradley D (ed.) *Papers in Southeast Asian linguistics: language policy, language planning and sociolinguistics in Southeast Asia*. Canberra, Australia: Pacific Linguistics. 131–154.
- Armstrong L E & Pe Maung Tin (1925). *A Burmese phonetic reader*. London: University of London Press.
- Bradley D (1982). 'Register in Burmese.' In Bradley D (ed.) *Pacific Linguistics Series A-62: Tonation*. Canberra, Australia: Pacific Linguistics, Australian National University.
- Bradley D (1995). *Papers in South Asian linguistics 13: Studies in Burmese linguistics*. Canberra, Australia: Pacific Linguistics, Australian National University.
- Myanmar Language Commission (1993). *Myanmar–English dictionary*. Yangon, Myanmar: Myanmar Language Commission.
- Myanmar Language Commission (2001). *English–Myanmar dictionary*. Yangon, Myanmar: Myanmar Language Commission.
- Okell J (1965). 'Nissaya Burmese, a case of systematic adaptation to a foreign grammar and syntax.' In Milner G B & Henderson E J A (eds.) *Indo-Pacific linguistic studies, vol. 2: Descriptive linguistics* (Lingua 14–15). Amsterdam: North Holland. 186–230.
- Okell J (1969). *A reference grammar of colloquial Burmese*. London: Oxford University Press.
- Okell J (1984). *Burmese: an introduction* (4 vols). DeKalb, IL: Northern Illinois University.
- Okell J & Allott A (2001). *Burmese/Myanmar: a dictionary of grammatical forms*. Richmond, UK: Curzon Press.
- Roop D H (1972). *An introduction to the Burmese writing system*. New Haven, CT: Yale University Press.
- Sprigg R K (1957). 'Studies in linguistics analysis.' *Transactions of the Philological Society* (Special volume). 104–138.
- Sprigg R K (1977). 'Tonal units and tonal classification: Panjabi, Tibetan and Burmese.' In Gill H S (ed.) *Pākha Sanjam 8: Parole and langue*. Patiala: Punjabi University. 1–21.
- Thurgood G W (1981). *Monumenta Serindica 9: Notes on the origins of Burmese creaky tone*. Tokyo: Tōkyō gaikokugo daigaku.
- Wheatley J K (1990). 'Burmese.' In Comrie B (ed.) *The major languages of East and Southeast Asia*. London: Routledge. 106–126.
- Wheatley J K (1996). 'Burmese writing.' In Daniels P T & Bright W (eds.) *The world's writing systems*. Oxford: Oxford University Press. 450–456.
- Wheatley J K (2003). 'Burmese.' In Thurgood G & LaPolla R J (eds.) *The Sino-Tibetan languages*. London & New York: Routledge. 195–207.

Relevant Websites

- Sino-Tibetan etymological dictionary and thesaurus (STEDT) (2002). University of California at Berkeley. <http://linguistics.berkeley.edu>.
- SIL International (2002). *Ethnologue: Languages of Myanmar*. <http://www.ethnologue.com>.

Burushaski

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

Burushaski is a language isolate spoken in the Northern Areas, Pakistan, primarily in the Hunza, Nagar, and, Yasin valleys. A small enclave of Burushaski speakers is also found over the border in Kashmir, India. The Hunza and Nagar varieties differ only minimally from each other; both stand at a relative distance from the Yasin variety of Burushaski, sometimes also considered to be a close sister language, Werchikwar.

There are approximately 80 000 speakers of Burushaski, including somewhere in the area of 15 000–20 000 people speaking the Yasin dialect,

with an additional 20 000–30 000 speakers of both Hunza Burushaski and Nagar Burushaski. In all communities where Burushaski is spoken, the language remains vital, with many women and children still monolingual speakers.

The first comprehensive study of Burushaski was Lorimer (1935–1938). The most recent is Berger's three-volume grammar, dictionary, and text collection (1998).

Bilingualism among Burushaski speakers is common primarily in the two Dardic Indo-European languages Shina (Nagar Burushaski speakers) and Khowar (the Burusho of Yasin valley). In Hunza, especially in the village of Mominabad, the Indo-Aryan-speaking Dúumaki (Domaaki) live in close contact with Burushaski speakers; nearly all Dúumaki speakers appear to be bilingual in Burushaski. Burushaski

itself may have previously been spoken in a wider area than it is currently found: for example, in Dras, in Baltistan, there is a group of people known as the Brokpa or Brusa; also, in Ponjal, there are the so-called Burushken, who are now Shina speaking.

Burushaski has a basic five-vowel system, with two series of contrastive long vowels, alternatively bearing stress or higher pitch on the first or second mora, respectively:

| | | | | | | | |
|-----|---|----|----|---|----|----|----|
| (1) | i | íi | íí | | u | úu | úú |
| | e | ée | eé | | o | óo | oó |
| | | | | a | áa | aá | |

There is some dispute among Burushaski specialists as to the exact nature of these long vowels. Varma (1941: 133) described the suprasegmental or intonational contrasts of Burushaski long vowels as representing a rising and falling tone; modern investigators, however, e.g., Tiffou (1993), Berger (1998), and Morin and Tiffou (1989), considered this to be a difference of moraic stress: that is, Burushaski long vowels may receive stress on either the first mora or the second, corresponding to Varma's falling and rising tones, respectively. These phenomena are phonemic in Burushaski. A comprehensive instrumental analysis of Burushaski vocalism remains to be done. A lowered pitch on the first mora is sometimes heard with the former (initial-mora prominent) forms. (Note that expressive diminutives are generally associated with this intonational pattern, e.g., *šon* 'blind' vs. *šoon* 'somewhat blind' or *tak* 'attached' vs. *táak* 'somewhat attached'.) Yasin exhibits the same intonational phenomena as the standard Hunza and Nagar varieties, although the moraic stress difference seems to be less pronounced, and in some speakers, this contrast has been neutralized.

Examples of phonemic vowel contrasts in Burushaski include *bat* 'flat stone' vs. *baát* 'porridge' (as in *bras-e baát* 'cooked rice,' *aalu-e baát* 'mashed potatoes'); *dir* 'boundary, water ditch between fields, small irrigation canal; hostility' vs. *đir* 'overhanging rock'; *yun* 'wooden block in door lock, stocks (for prisoner)' vs. *yúun* 'quail'; *men* 'who' vs. *meén* 'old, venerable; fallow field'; *gon* 'dawn' vs. *goón* 'like, as.' Note that these length contrasts only appear in stressed syllables in Burushaski.

Three-way contrasts between short, first-mora-prominent, and second-mora-prominent vowels are found in a small number of lexical items in Burushaski. Such triplets include *bo* 'grain, seed, sperm/semen' vs. *bóo et-* 'low, bellow' vs. *boó* (cf. *nupáu* ~ *nupoón* in the converb form) 'sit down, lower self,' *don* 'large herd' vs. *dóon* (~*dóon ke*) 'still, yet, nevertheless' vs. *doón* 'woman's head scarf; open' (Berger, 1998: vol. 3, pp. 121–122). Two-way length contrasts, such as

báak 'punishment, torture' vs. *baák* 'generosity' are relatively common.

Burushaski has an extensive system of consonants. In fact, there are eight different stop/affricate series attested in the language. This includes labial, dental, alveolar, retroflex, palatal, palatal-retroflex, velar, and uvular. All of these series may be found in voiceless unaspirated, voiceless aspirated, and voiced series (see Table 1).

While retroflexion is common throughout the languages of south Asia, Burushaski has one of the largest inventories of nonsonorant retroflex sounds among the languages of the region, with no fewer than seven such sounds. In addition, the Hunza and Nagar varieties possess a curious retroflex, a spirantized palatal, symbolized /ɣ/, with a range of local or idiolectal realizations. This sound is lacking in the Yasin Burushaski dialect.

Burushaski possesses four noun classes, based on real-world semantic categorization. Thus, male humans belong to class I, female humans to class II, nonhuman animates to class III and inanimates to class IV (2). These classes are formally realized not in the noun themselves but through the selection of case allomorphs and verb agreement morphology.

| | | |
|-----|-----------------------|-----------------------|
| (2) | I: male human | II: female human |
| | <i>bir</i> 'man' | <i>dasín</i> 'girl' |
| | III: animate nonhuman | IV: inanimate |
| | <i>hayúr</i> 'horse' | <i>yaténč</i> 'sword' |

Another salient feature of the nominal system of Burushaski is the wide range of plural formations attested in the language. There are literally dozens of plural markers in the language, each often found with only a small number of nouns. Sometimes these are found only with nouns of a particular class but others crosscut this categorization (see Table 2).

Burushaski has a highly developed system of grammatical and instrumental cases as well as an elaborate system of local/directional cases and instrumental/comitative cases (see Table 3). The exact number is difficult to determine as new elements enter this

Table 1 The consonantal inventory of Burushaski

| | | | | | | | |
|------------------|----------------|----------------|-----------------|----------------|-----------------|------------------|----------------|
| p | t | c | t̪ | č | č̪ | k | q |
| p ^h | t ^h | c ^h | t̪ ^h | č ^h | č̪ ^h | k ^h | q ^h |
| b | d | z | d̪ | ǰ | ǰ̪ | g | ɣ |
| (f) ^a | | s | | š | š̪ | (x) ^a | h |
| m | n | | | | | ŋ | |
| w | | | | ɣ | ɣ̪ | | |
| | l | r | | | | | |

^a[f] and [x] occur only in loan words, or as a variant of the aspirated stops [p^h] and [q^h] or [k^h], respectively.

system through the grammaticalization (and phonological fusion) of relational nouns/postpositions. There are at least the following grammatical cases (i.e., ones assigned by structural position or verbal subcategorization): ergative, genitive, dative, ablative. In the latter two instances with class II nouns, the cases are built off the genitive (or oblique) stem.

Table 2 Plural formation in Burushaski

| Singular | Plural | |
|----------------|----------------------------|-----------------|
| <i>hal</i> | <i>hal-jó</i> | 'fox' |
| <i>jiip</i> | <i>jiip-uc</i> | 'jeep' |
| <i>γus</i> | <i>γuš-oŋo</i> | 'earthen clump' |
| <i>čonč</i> | <i>čonč-iŋ</i> | 'summit, peak' |
| <i>-yarum</i> | <i>γarum-iŋ ~ γarim-iŋ</i> | 'part' |
| <i>girkis</i> | <i>girkič-o</i> | 'rat' |
| <i>γurkun</i> | <i>γurkuy-o</i> | 'frog' |
| <i>γurkuc</i> | <i>γurkuč-o</i> | 'frog' (Nagar) |
| <i>ašaáto</i> | <i>ašaátu-tiŋ</i> | 'weak(ling)' |
| <i>γat-enč</i> | <i>γat-aŋ</i> | 'sword' |

Numerals agree in class with their nominal complement in class in Burushaski (note class-I and class-III are conflated here; see **Table 4**). Numbers 20 and above are based on a clear vigesimal system, 30 literally being '20+10' and 40 being (etymologically) '2+20.' etc.

- (3) *aalter(an)* 20 *aalter toorumo* 30
aaltuwalter 40 *aaltuwalter toorumo* 50
iiski aalter 60 *iiski aalter toorumo* 70
waalti aalter(an) 80 *waalti aalter toorumo* 90 *t^ba* 100

The verbal system of Burushaski stands out for its morphological complexity among south Asian languages. There are two basic sets of inflections, depending in part on the stem allomorph. These two broad categories are as follows:

- | | |
|-------------------|-----------|
| (4) I | II |
| past | future |
| perfect | present |
| pluperfect | imperfect |
| aorist (conative) | |

Table 3 Case forms in Burushaski

| | 'man' [I] | 'woman' [II] | 'horse' [III] | 'sword' [IV] |
|--------------------------------------|---------------------|------------------|-------------------|-------------------|
| Grammatical cases | | | | |
| NOM/ABS | <i>hir</i> | <i>gus</i> | <i>haγur</i> | <i>γatenč</i> |
| ERG | <i>hir-e</i> | <i>gus-e</i> | <i>haγur-e</i> | <i>γatenč-e</i> |
| GEN | <i>hir-e</i> | <i>gus-mu</i> | <i>haγur-e</i> | <i>γatenč-e</i> |
| OBLQ.stem | <i>hir-</i> | <i>gusmu-</i> | <i>haγur-</i> | <i>γatenč-</i> |
| DAT | <i>hir-ar</i> | <i>gusmu-r</i> | <i>haγur-ar</i> | <i>γatenč-ar</i> |
| ABL | <i>hir-cum</i> | <i>gusmu-cum</i> | <i>haγur-cum</i> | <i>γatenč-cum</i> |
| Local-Directional Cases | | | | |
| <i>gus-mu-te</i> | | | | |
| woman-II.OBLQ-SUPERESS | | | | |
| 'on the woman' | | | | |
| <i>ǰakun</i> | <i>un-ale</i> | <i>bi-m</i> | | |
| donkey | you-ADESS | be-III.AP | | |
| 'the donkey was near you' | | | | |
| <i>e-š-aŋum</i> | | | | |
| I-neck-SUPERABL | | | | |
| 'from on his neck' | | | | |
| Instrumental/Comitative Cases | | | | |
| <i>uskó</i> | <i>yát-umuc-aŋe</i> | <i>hin</i> | <i>jinzaat-an</i> | |
| three | head-PL-INSTR.B | one.I | demon-SG.ART | |
| 'a three-headed demon' | | | | |
| <i>day-o-k</i> | <i>d-l</i> | | | |
| stone-PL-INSTR | hit | | | |
| 'pelt with stones' | | | | |
| <i>-me-ke</i> | <i>gał</i> | | | |
| tooth-INSTR | bite | | | |
| 'bite with teeth' | | | | |
| <i>ǰmé-k</i> | <i>d-l</i> | | | |
| bow-INSTR | hit | | | |
| 'shoot with bow' | | | | |
| <i>ǰamé-k-aŋe</i> | <i>bišá-</i> | | | |
| bow-INSTR-SUPERESS | throw | | | |
| 'shoot with bow' | | | | |

Table 4 Numerals

| | I/III | II | IV |
|----|------------------|-------------------|--------------------|
| 1 | <i>hin</i> | <i>han</i> | <i>hi(k)</i> |
| 2 | <i>aaltan</i> | <i>aala/aalto</i> | <i>aalti/aalto</i> |
| 3 | <i>iisken</i> | <i>usko</i> | <i>iiski</i> |
| 4 | <i>waalto</i> | <i>waalto</i> | <i>waal(ti)</i> |
| 5 | <i>cundo</i> | <i>cundo</i> | <i>cindi</i> |
| 6 | <i>mišindo</i> | <i>mišindo</i> | <i>mišin(di)</i> |
| 7 | <i>talo</i> | <i>talo</i> | <i>tale</i> |
| 8 | <i>aaltambo</i> | <i>aaltambo</i> | <i>aaltam(bi)</i> |
| 9 | <i>hunčo</i> | <i>hunčo</i> | <i>hunti</i> |
| 10 | <i>toorumo</i> | <i>toorumo</i> | <i>toorimi</i> |
| 11 | <i>turma hin</i> | <i>turma han</i> | <i>turma hik</i> |

The maximal template of the Burushaski simplex verb is given by Tikkanen (1995: 91) as:

| | | | | | |
|-----|-------------|------|----------------------|-----------|----|
| (5) | NEG- | D- | PERSON/CLASS/NUMBER- | CAUS- | ✓- |
| | -4 | -3 | -2 | -1 | Ø |
| | PL.SUBJ- | DUR- | | 1SG.SUBJ- | |
| | +1 | +2 | | +3 | |
| | PRTCPL/OPT/ | | SUBJ.SFX- | | Q |
| | COND/AUX- | | | | |
| | +4 | | +5 | | +6 |

Some examples of verbs reflecting this template are given in (6). Note the curious and morphologically triggered (and phonologically unmotivated) devoicing of obstruents following the negative allomorph *a-* (but not *oó-*).

| | | |
|-----|---------------------------|----------------------------------|
| (6) | <i>oó-min-im-i</i> | <i>oó-man-um-an</i> |
| | NEG-drink-AP-I | NEG.PL-become-AP-PL |
| | 'he didn't drink (it)' | 'they didn't become' |
| | (Berger, 1998: 106) | (Berger, 1998: 106) |
| | <i>a-túru-m-i</i> | <i>duróo-m-i</i> |
| | NEG-work-AP-I | work-AP-I |
| | 'he didn't work' | 'he worked' |
| | (Berger, 1998: 105) | (Berger, 1998: 105) |
| | <i>a-mí-kač-ič-a-i</i> | <i>mi-k^báč-ič-a-i</i> |
| | NEG-1PL-enclose-DUR-AUX-I | 1PL-enclose-DUR-AUX-I |
| | 'he doesn't enclose us' | 'he encloses us' |
| | (Berger, 1998: 105) | (Berger, 1998: 105) |
| | <i>a-tu-ququ-m-i</i> | <i>du-q^bóqu-m-i</i> |
| | NEG-D-be.confused-AP-I | D-be.confused-AP-I |
| | 'he was not confused' | 'he was confused' |
| | (Berger, 1998: 105) | (Berger, 1998: 105) |

In addition to subject and direct/indirect objects, the Burushaski verb may also optionally encode an animate possessor of a logical argument as an argument morphologically in the verb-word (7).

| | | | |
|------|--|-------------------------|-----------------|
| (7a) | <i>k^bak^báy-umuc</i> | <i>p^bašú</i> | <i>mée-t-aa</i> |
| | walnut-PL | gobble.up | 1PL-AUX-2 |
| | 'you gobbled up our walnuts' (Berger, 1998: 162) | | |

| | | | | |
|------|--|-----------------|---------------|-----------------------|
| (7b) | <i>hiles-e</i> | <i>dasin-mo</i> | <i>mo-miš</i> | <i>moo-skarc-im-i</i> |
| | boy-ERG | girl-GEN | II-finger | II-cut-AP-I |
| | 'the boy cut off the girl's finger' (Willson, 1990: 5) | | | |

Another characteristic feature of the Burushaski verbal system is the grammaticalized use of double argument indexing with intransitive verbs. This single vs. double marking appears within two separate functional subsystems. In the first one, presence vs. absence of double marking implies degree of control of the subject over the action: less control is indexed through double marking (8a). In the second such subsystem, class-IV nouns receive single marking while class-III nouns receive double marking with the same predicate (8b).

(8a) *yurc-im-i*
sink-AP-I
'he dove under' (Berger, 1998: 118)
i-yúrc-im-i
I-sink-AP-I
'he drowned' (Berger, 1998: 118)

(8b) *ha* *yulú-m-i*
house burn-AP-IV
'the house burned' (Berger, 1998: 118)
hun *i-yúl-im-i*
wood III-burn-AP-III
'the wood burned' (Berger, 1998: 118)

Syntactically, Burushaski is a fairly rigid SOV language. In narrative texts, head-tail linkage, a common narrative device among south Asian languages, is frequently found (clauses are linked by rote repetition of the finite verb of a preceding sentence in a nonfinite form in an immediately following sentence). Further, some cases appear only on the leftmost of two (conjunctively or disjunctively) conjoined nouns, while others appear on both. There thus appear to be both phrasal and word-level case forms in Burushaski. A further curious aspect of Yasin Burushaski is the highly atypical semantic (plural) agreement seen with disjunctively conjoined NPs (Anderson and Eggert, 2001). Most of these features can be seen in the following examples.

| | | | | | |
|------|--|-----------|---------------|--------------|---------------------|
| (9a) | <i>gus</i> | <i>ya</i> | <i>hir-e</i> | <i>dasen</i> | <i>a-mu-yeec-en</i> |
| | woman | or | man-ERG | girl | NEG-II-see-PL |
| | 'the woman or the man didn't see the girl' | | | | |
| | (Anderson <i>et al.</i> , 1998) | | | | |
| (9b) | <i>hir</i> | <i>ya</i> | <i>guse-e</i> | <i>dasen</i> | <i>a-mu-yeec-en</i> |
| | man | or | woman-ERG | girl | NEG-II-see-PL |
| | 'the man or the woman didn't see the girl' | | | | |
| | (Anderson <i>et al.</i> , 1998) | | | | |

Another characteristic feature of Burushaski syntax is the extensive use of case forms to mark a wide range of subordinate clause functions (Anderson, 2002).

- (10) *ma ma-ír-áte ĩe taŋ*
 y'all 2PL-die-SUPERESS I sad
a-máy-a-m
 1-become.dur-1-AP
 'when you all die I will be sad'
 (Berger, 1998: 140)

Burushaski includes loans from a range of local languages including Urdu, Khowar, Shina, and even (perhaps indirectly) from Turkic languages as well. In some instances, loan affixes may be found as well, e.g., *daḍaŋ-ci* 'big-drum drummer' (Berger, 1998: 209). More tenuous lexical connections have been proposed with Northeast Caucasian languages and Paleo-Balkan Indo-European languages (Casule, 1998).

There is a small body of indigenous literature in Burushaski written in a modified Urdu script. In addition, various texts in transcription have appeared, including Skyhawk *et al.* (1996), Skyhawk (2003), etc.

Bibliography

- Anderson G D S (1997). 'Burushaski phonology.' In Kaye A S & Daniels P T (eds.) *Phonologies of Asia and Africa (including the Caucasus)*. Winona Lake, IN: Eisenbrauns. 1021–1041.
- Anderson G D S (2002). 'Case marked clausal subordination in Burushaski complex sentence structure.' *Studies in Language* 26(3), 547–571.
- Anderson G D S & Eggert R H (2001). 'A typology of verb agreement in Burushaski.' *Linguistics of the Tibeto-Burman Area* 24(2), 235–254.
- Anderson G D S, Eggert R H, Zide N H & Ramat F (1998). *Burushaski language materials*. Chicago: University of Chicago Language Laboratories and Archives.
- Bashir E (1985). 'Towards a semantics of the Burushaski verb.' In Zide A, Magier R K D & Schiller E (eds.) *Proceedings of the Conference on Participant Roles: South Asia and Adjacent Areas*. Bloomington: Indiana University Linguistics Club. 1–32.
- Benveniste E (1949). 'Remarques sur la classification nominale en Burushaski.' *Bulletin de la Société Linguistique de Paris* 44, 64–71.
- Berger H (1956). 'Mittelmeerische Kulturpflanzennamen aus dem Burushaski.' *Münchener Studien zur Sprachwissenschaft* 9, 4–33.
- Berger H (1959). 'Die Burushaski-Lehnwörter in der Zigeunersprache.' *Indo-Iranian Journal* 3, 17–43.
- Berger H (1974). *Das Yasin-Burushaski (Werchikwar): Grammatik, Texte, Wörterbuch*. Wiesbaden: Otto Harrassowitz.
- Berger H (1994). 'Kombinatorischer Lautwandel im Burushaski.' *Studien zur Indologie und Iranistik* 19, 1–9.
- Berger H (1998). *Die Burushaski-Sprache von Hunza und Nager* (3 vols). Wiesbaden: Otto Harrassowitz.
- Bleichsteiner R (1930). 'Die werschikisch-burischkische Sprache im Pamir-Gebiet und ihre Stellung zu den Japhetitisprachen des Kaukasus.' *Wiener Beiträge zur Kulturgeschichte und Linguistik* 1, 289–331.
- Casule I (1998). *Basic Burushaski etymologies: the Indo-European and Paleo-Balkan affinities of Burushaski*. Munich: Lincom Europa.
- Klimov G A & Edel'man D I (1970). *Iazyk burushaski*. Moscow: Akademia Nauk SSSR.
- Leitner G W (1889). *The Hunza and Nagyr hand-book: being an introduction to a knowledge of the language, race, and countries of Hunza, Nagyr, and a part of Yasin*. Calcutta.
- Lorimer D L R (1932). 'A Burushaski text from Hunza.' *Bulletin of the School of Oriental Studies* 4, 505–531.
- Lorimer D A (1935–1938). *The Burushaski language* (3 vols). Oslo: H. Aschehoug.
- Morgenstierne G (1945). 'Notes on Burushaski phonology.' *Norsk Tidsskrift for Sprogvidenskap* 13, 59–95.
- Morgenstierne G, Vogt H & Borström C J (1945). 'A triplet of Burushaski studies.' *Norsk Tidsskrift for Sprogvidenskap* 13, 61–147.
- Morin Y-C & Tiffou E (1988). 'Passive in Burushaski.' In Shibatani M (ed.) *Passive and voice*. Amsterdam: John Benjamins. 493–525.
- Morin Y-C & Tiffou E (1989). *Dictionnaire complémentaire du Bourouchaski du Yasin*. Paris: Peeters/SELAF.
- Skyhawk H van (2003). *Burushaski-Texte aus Hispar: Materialien zum Verständnis einer archaischen Bergkultur in Nordpakistan*. Wiesbaden: Otto Harrassowitz.
- Skyhawk H van, Berger H & Jettmar K (1996). *Libi Kisar: ein Volksepos im Burushaski von Nager*. Wiesbaden: Otto Harrassowitz.
- Tiffou E (1977). 'L'Effacement de l'ergatif en bourouchaski.' *Studia Linguistica* 31, 18–37.
- Tiffou E (1993). *Hunza proverbs*. Calgary: University of Calgary Press.
- Tiffou E & Patry R (1995). 'La Notion de pluralité verbale: le cas du bourouchaski du Yasin.' *Journal Asiatique* 283(2), 407–444.
- Tiffou E & Pesot J (1988). *Contes du Yasin*. Paris: Peeters.
- Tikkanen B (1995). 'Burushaski converbs in their areal context.' In Haspelmath M & König E (eds.) *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds*. Berlin: Mouton de Gruyter. 487–528.
- Toporov N V (1970). 'About the phonological typology of Burushaski.' In Jakobson R & Kawamoto S (eds.) *Studies in general and Oriental linguistics presented to Shiro Hattori on the occasion of his sixtieth birthday*. Tokyo: TEC Corporation for Language and Educational Research. 632–647.
- Toporov V N (1971). 'Burushaski and Yeniseian languages: some parallels.' In van Poldauf I (ed.) *Etudes de la phonologie, typologie et de la linguistique générale*. Prague: Académie Tchécoslovaque des Sciences. 107–125.
- Varma S (1941). 'Studies in Burushaski dialectology.' *Journal of the Royal Asiatic Society of Bengal Letters* 7, 133–173.
- Willson S R (1990). Verb agreement and case marking in Burushaski. M.A. thesis, University of North Dakota.

This page intentionally left blank

C

Caddoan Languages

D Rood, University of Colorado, Boulder,
CO, USA

© 2006 Elsevier Ltd. All rights reserved.

Caddoan is a family of North American language consisting of two branches: Caddo, formerly spoken in Texas and Louisiana, and now spoken only in Oklahoma; and North Caddoan, found in the central Plains from Oklahoma to North Dakota. The North Caddoan languages include Arikara, Pawnee, Kitsai, and Wichita. Arikara and Pawnee are linguistically very close, while Kitsai falls between them and Wichita.

Language Structure

The Caddoan languages have extremely small phoneme inventories, but complex morphophonemics. They are morphologically and syntactically prototypical examples of polysynthetic structure. The proposed phoneme inventory for the family is */p, t, k, c (= [ts]), s, w, n, r, y, ʔ, h, i, a, u/ (Chafe, 1979: 218–219). Caddo has a somewhat larger set, which appears to result from relatively recent expansion.

Caddoan verbs consist of 30 or more positional slots into which bound morphemes may be inserted; the verb root occurs near the end. In addition to expected categories like tense, modality, aspect, pronoun, number, evidential, and verb root, there are slots for certain adverbs, incorporated objects, patient definiteness (in Wichita and possibly others), and derivational stem-forming elements. All the languages have a bipartite verb stem for many verbs; a class of ‘preverbs’ occurs separated from the root by several slots.

Nouns generally may take only one of two or three suffixes: an ‘absolutive’ (which occurs only when the noun is used alone), a locative, or, in some of the languages, an instrumental. Noun compounds are frequent and productively formed. All the languages lack adpositions and most adjectives.

Sentential argument structure (subject, object, indirect object, possessor) is marked entirely in the

verbal complex; word order in clauses has strictly pragmatic functions. Intransitive verbs fall into two classes depending on whether their subjects are marked by transitive object pronouns or transitive agent pronouns.

History and Scholarship

Europeans first encountered speakers of Caddoan languages during the 16th-century Spanish expeditions from Mexico searching for Quivira (the land supposed to have included El Dorado, a rumored but non-existent city with streets of gold). Maps from those expeditions record a few (now largely uninterpretable) place names, but beyond that most information on the languages has been collected since the 1960s. Kitsai was recorded as spoken by its last monolingual speaker in the early 20th century, but none of the data has been published. The other languages continued to have a few speakers at the beginning of the 21st century, but all will probably be extinct by 2025, despite language preservation and revival efforts.

Large text collections and good grammars are available for two of the languages, Arikara and Pawnee, thanks to the work of Douglas R. Parks. Parks has also coauthored a series of Arikara teaching grammars and a dictionary for elementary school students. Wichita is documented in a grammar, several articles about grammatical phenomena, and a few texts by David S. Rood, as well as audio and video documentation archived at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. For Caddo, see the texts by Wallace L. Chafe and the detailed description of verb morphology by Lynette Melnar. Allan R. Taylor and W. L. Chafe have published on the history of the Caddoan language family (see Chafe, 1979, for further reading).

Bibliography

Chafe W L (1979). ‘Caddoan.’ In Campbell L & Mithun M (eds.) *The languages of native America: Historical and comparative assessment*. Austin, TX: University of Texas Press.

- Chafe W L (2005). 'Caddo.' In Hardy H K & Scancarelli J (eds.) *The native languages of the southeastern United States*. Lincoln, NE: University of Nebraska Press.
- Melnar L R (2004). *Caddo verb morphology*. Lincoln, NE: University of Nebraska Press.
- Parks D R (1976). *A grammar of Pawnee*. New York: Garland.
- Parks D R (ed.) (1977). *Native American texts series*, vol. 2, no. 1: *Caddoan texts*. Chicago: University of Chicago Press.
- Parks D R (1991). *Traditional narratives of the Arikara Indians* (4 vols). Lincoln, NE: University of Nebraska Press.
- Parks D R (2005). *An elementary dictionary of Skiri Pawnee*. Lincoln, NE: University of Nebraska Press.
- Parks D R, Beltran J & Waters E P (1998–2001). *An introduction to the Arikara language: Sabni¹ Wakuunu* (2 vols). Roseglen, ND: White Shield School. [Multi-media versions on CD are available from the American Indian Research Institute, Bloomington, IN.]
- Rood D S (1976). *Wichita grammar*. New York: Garland.
- Rood D S & Lamar D J (1992). *Wichita language lessons* (manual and tape recordings). Anadarko, OK: Wichita and Affiliated Tribes.

Cape Verdean Creole

M Baptista, University of Georgia, Athens, GA, USA

© 2006 Elsevier Ltd. All rights reserved.

Cape Verdean Creole (henceforth CVC) is spoken in Cape Verde Islands, an archipelago located in the Atlantic Ocean off the northwestern coast of Africa, at approximately 450 kilometers from Senegal. The archipelago is divided into two main clusters: the windward islands (locally known as Barlavento) and the leeward islands (Sotavento). Barlavento includes Boavista, Sal, São Nicolau, Santa Luzia, São Vicente, and Santo Antão. Sotavento consists of Brava, Fogo, Santiago, and Maio.

Given the strategic location of the archipelago at the crossroads of Europe, Africa, and America, the Portuguese settled the islands from 1462 onward, and the islands came to play a critical role in the slave trade from the 15th to the 19th centuries. As a result, many view CVC as the oldest creole alive today. Historical sources (Brásio, 1962) state that the tribes of Mandingues, Balantes, Bijagos, Feloupes, Beafadas, Pepels, Quissis, Brames, Banhuns, Peuls, Jalofos, Bambaras, Bololas, and Manjakus provided most of the human contingent to the slave trade in Cape Verde. The white settlers came from Algarve and Alentejo in Portugal and also included Jews, Spaniards, Italians, and French (Martinus, 1996). Having been settled at different times with different populations, it is not surprising that a number of morphophonological and syntactic features distinguish Barlavento varieties (closer to Portuguese) from their Sotavento counterparts (more Africanized), resulting in a fairly complex sociolinguistic situation.

Although earlier descriptions of the language viewed CVC as a mere dialect of Portuguese, recent studies have shed new light on the hybrid nature of CVC focusing on the African contributions to the formation of the language. Baptista (2003a) studied specifically reduplication, a morphological process found in African languages whereby a reduplicated adjective or adverb expresses emphasis, as in *moku moku* 'very drunk' or *faxi faxi* 'very quickly'. Noun reduplication may yield a distributive interpretation, as in *dia dia* 'every day' or may simply lead to a change in meaning, as in *boka* 'mouth,' *boka boka* signifying 'in secret'. Lexical categories such as adjectives once reduplicated may shift category (i.e., adjective to noun) as in *mansu* 'quiet', *mansu mansu* 'secrecy'. Other scholars such as Rougé (2004) and Quint (2000) have examined the possible African etymology of some of the Cape Verdean linguistic items that have found their way in the grammatical and lexical components of the language. Lang (2004) has investigated how some grammatical morphemes inherited from Portuguese may also take on new functions passed down from substrates like Wolof. In a similar vein of work, Baptista (2003b) has examined how the plural suffix-s in Cape Verdean inherited from Portuguese is sensitive to conditions such as the animacy hierarchy and definiteness, two variables playing a role in the African languages having contributed to the genesis of CVC.

Such studies demonstrate the genuine hybrid nature of CVC by examining how various elements from all source languages involved in its genesis interact and at what level. This gives us valuable insights into cognitive processes at play when languages come abruptly into contact.

Bibliography

- Baptista M (2002). *The syntax of Cape Verdean Creole: the Sotavento varieties*. Amsterdam/Philadelphia: John Benjamins.
- Baptista M (2003a). 'Number inflection in creole languages.' *Interface* 6, 3–26.
- Baptista M (2003b). 'Reduplication in Cape Verdean Creole.' In Kouwenberg S (ed.) *Twice as meaningful: reduplication in pidgins and creoles*. London: Battlebridge. 177–184.
- Brásio A (1962). 'Descobrimento, povoamento, evangelização do archipelago de Cabo Verde.' *Studia* 10, 49–97.

- Lang J (2004). *Dicionário do crioulo da ilha de Santiago (Cabo Verde)*. Tübingen: Gunter Narr Verlag.
- Martinus F (1996). *The kiss of a slave: Papiamentu's West-African connections*. Ph.D. diss., University of Amsterdam.
- Quint N (2000). *Grammaire de la langue cap-verdienne*. Paris: L'Harmattan.
- Rougé J L (2004). *Dictionnaire étymologique des créoles portugais d'Afrique*. Paris: Karthala.
- Veiga M (1998). *Le Créole du Cap-Vert: étude grammaticale descriptive et contrastive*. Ph.D. diss., Université Aix-Marseille.

Cariban Languages

S Meira, Leiden University, Leiden, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The Cariban family is one of the largest genetic groups in South America, with more than 25 languages (see **Figure 1**) spoken mostly north of the Amazon, from Colombia to the Guianas and from northern Venezuela to Central Brazil (see **Figure 2**). Despite the long history of their studies, most Cariban languages are still insufficiently described. The best descriptive works published so far are Hoff (1968, on Karinya) and Derbyshire (1979, 1985, on Hishkaryana). There are good descriptive works on Apalai, Makushi, and Waiwai in Derbyshire and Pullum (1986–1998); Jackson (1972) gives a brief, but detailed, overview of Wayana. Muller (1994) is a very informative Panare dictionary. Meira (2005) and Carlin (2004) are full descriptions of Tiriyo; Meira (2000), mostly a historical study, contains some descriptive work on Tiriyo, Akuriyo, and Karihona. Gildea (1998) and Derbyshire (1999) contain surveys of the family.

Comparative Studies and Classification

First recognized by the Jesuit priest Filippo Salvatore Gilij in the 18th century (Gilij, 1780–1783), the Cariban family was subsequently studied by L. Adam (1893) and C. H. de Goeje (1909, 1946). After some initial tentative proposals within larger South American classifications (the last of which is Loukotka, 1968), the first detailed classification was published by V. Girard (1971), followed by M. Durbin (1977) and T. Kaufman (1994). Durbin's classification – unfortunately used in the Ethnologue (SIL) – is, as

Gildea (1998) pointed out, seriously flawed; Girard's classification is limited (14 low-level subgroups); Kafuman's classification is probably the best; it is based not on firsthand sources but on the comparison of other classifications. The proposal in **Figure 1** is the preliminary result of ongoing comparative research. There is some good evidence that Cariban and Tupian languages are distantly related (Rodrigues, 1985); other hypotheses (e.g., Ge-Pano-Carib and Macro-Carib, from Greenberg, 1987) remain mostly unsupported and are not accepted by specialists.

Shafer (1963) was the first attempt at reconstructing Proto-Cariban phonology, but its many flaws make Girard (1971) the real first proposal in this area. The most up-to-date study is Meira and Franchetto (2005). Meira (2000) reconstructs the phonology and morphology of the intermediate proto-language of the Taranoan subgroup.

Main Linguistic Features

Phonology

Cariban languages have small segmental inventories: usually only voiceless stops (*p*, *t*, *k*, *ʔ*), one or two fricatives/affricates (*h* or *ʕ*, *s* or *ʃ* or *tʃ*), two nasals (*m*, *n*), a vibrant (*r*, often *ɹ* or *ʀ*), glides (*w*, *j*), and six vowels (*a*, *e*, *i*, *o*, *u*, *ɨ*). Some languages have distinctive voiced obstruents (Bakairi, Ikpeng, Karihona), more than one vibrant or lateral (Bakairi, Kuikuro, Ikpeng, Hishkaryana, Waiwai, Kashuyana), or more fricatives or affricates (Bakairi, Waimiri-Atroari, Kashuyana, Waiwai); others have an extra vowel *ə* (Wayana, Tiriyo, Panare, Bakairi, Pemong, Kapong). Vowel length is often distinctive, whereas nasality usually is not, with few exceptions (Apalai, Bakairi, Kuikuro).

| | | |
|---|--|--------|
| 1. GUIANAN | | |
| a. | Karinya (Carib, Galibi, Kali'ña) | 10 000 |
| | Tiriyo (Trio) | 2000 |
| b. <i>Taranoan</i> : | Akuriyo (Akurio, Wama, Oarikule) | 3-4 |
| | Karihona (Carijona) | 5-10 |
| c. | Wayana (Roucouyenne, Urucuyana) | 750 |
| d. | Apalai (?) | 450 |
| e. | Palmella (†) (?) | |
| f. <i>Parukotoan</i> : | Waiwai | 1000 |
| | Hishkaryana (Hixkaryana) | 550 |
| | Kashuyana (Katzuyana, Xikuyana, Kahyana) | 90 |
| 2. VENEZUELAN | | |
| a. <i>Coastal</i> : | Chayma (†) | |
| | Kumanakoto (Cumanagota) (†) | |
| b. | Tamanaku (†) | |
| | Pemong | 5000 |
| c. <i>Pemongan</i> or <i>Pemong</i> <i>Proper</i> : | Arekuna, Kamarakoto, Taurepang | |
| | Kapong | 5000 |
| | Akawaio, Patamuna, Ingarikó | |
| | Makushi (Macuxi, Macushi) | 11 400 |
| d. | Panare | 1200 |
| e. | De'kwana (Ye'kwana, Maquiritare, Maiongong) (?)..... | 5000 |
| f. | Mapoyo (?) | 2 |
| g. | Yawarana (Yabarana) (?) | 20 |
| 3. WAIMIRIAN | | |
| a. | Waimiri-Atroari | 1000 |
| | Waimiri, Atroari | |
| 4. YUKPAN | | |
| a. | Yukpa (Motilón) | 3000 |
| b. | Hapreria (Japreria) | 80 |
| 5. PEKODIAN | | |
| a. <i>Xinguan</i> : | Ikpeng | 350 |
| | Arara | 200 |
| b. | Bakairi | 900 |
| | Eastern Bakairi, Western Bakairi | |
| 6. KUIKUROAN | | |
| a. | Kuikuro | 900 |
| | Kuikuro, Kalapalo, Nahukwa, Matipu | |
| b. | Pimenteira (?) (†) | |

Figure 1 A tentative classification of Cariban languages. (?) = difficult to classify; (†) = extinct (not all listed here). Different names or spellings for the same language are given in parentheses. Dialects are indented under the language name. (Demographic data refer to speakers, not ethnic members of the group; sources: Ethnologue and author's own work).

Many languages have weight-sensitive rhythmic (iambic) stress (Table 1; Meira, 1998); some, however, have simple cumulative, usually penultimate, stress (Panare, Bakairi, Kuikuro, Yukpa). Morphophonological phenomena include stem-initial ablaut in verbs and nouns and the systematic reduction of stem-final syllables within paradigms (Gildea, 1995; Meira, 1999).

Morphology

Cariban languages are mostly suffixal; prefixes exist also, marking person and valency (the latter on verbs). Some languages (Tiriyo, Wayana, Apalai) have reduplication. The complexity of the morphology is comparable to that of Romance languages. There are usually nouns, verbs, postpositions, adverbs (a class that includes most adjectival notions), and particles.

Possessed nouns take possession-marking suffixes that define subclasses (*-ri*, *-ti*, *-ni*, *-Ø*) and person-marking prefixes that indicate the possessor (e.g., Ikpeng *o-megum-ri* 'your wrist', *o-muj-n* 'your boat', *o-eg-Ø* 'your pet'). With overt nominal possessors, some languages have a linking morpheme *j-* (e.g., Panare *Toman j-uwə?* 'Tom's house, place'). Nouns can also be marked for past ('ex-N', 'no longer N') with special suffixes (*-tpo*, *-tpi*, *-bi*, *-tpə*, *-hpə*, *-npə*, etc.; e.g., Bakairi *ūwī-bi-ri* 'my late father'). Pronouns distinguish five persons (1, 2, 3, 1 + 2 = dual inclusive = 'you and I', 1 + 3 = exclusive; the 1 + 3 pronoun functions syntactically as a third-person form) and two numbers (singular, or noncollective, and plural, or collective). The third-person forms also have gender (animate vs. inanimate) and several deictic distinctions (Table 2). To each pronoun usually

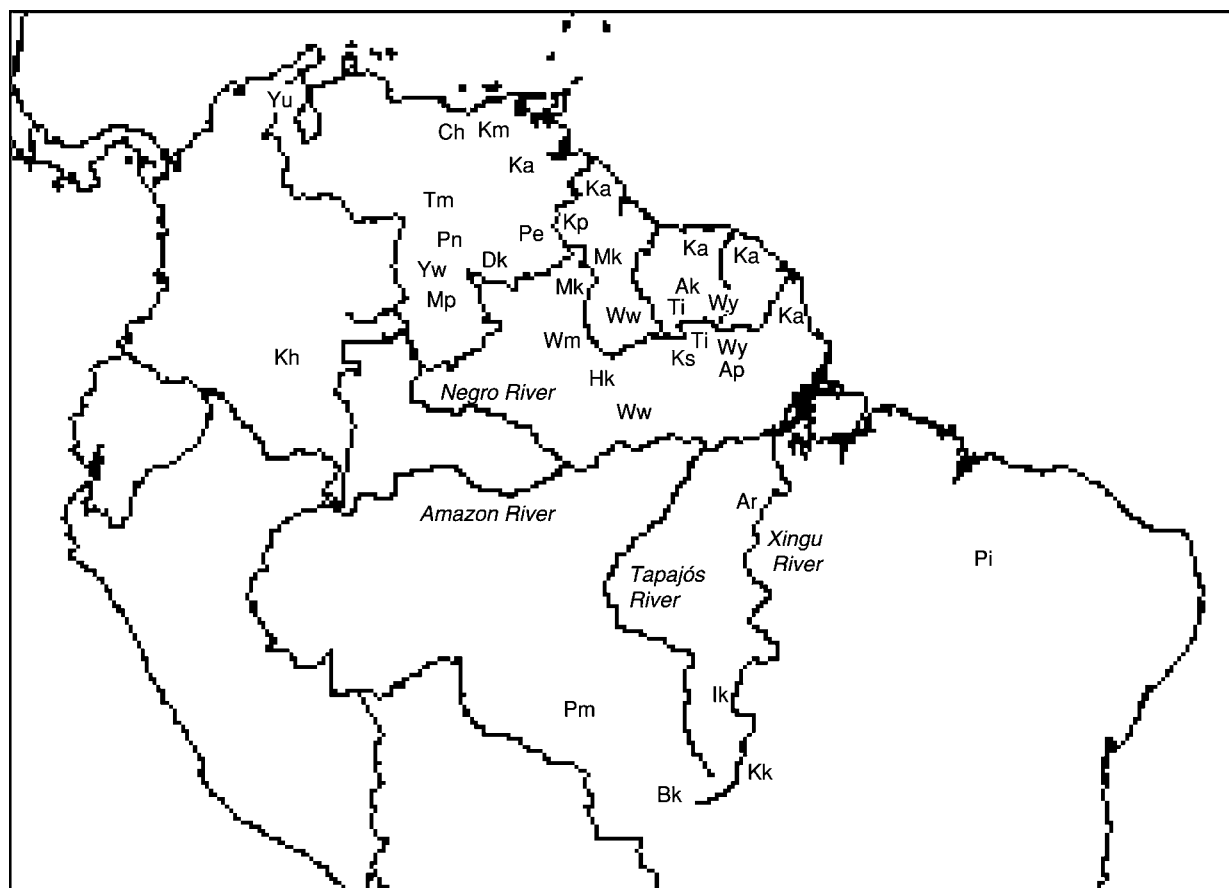


Figure 2 Map of the current distribution of Cariban languages. Living languages in bold, extinct languages in normal type. **AK**, Akuriyo; **Ar**, Arara; **Bk**, Bakairi; **Ch**, Chayma; **Dk**, De'kwana; **Hk**, Hishkaryana; **Ik**, Ilkpeng; **Ka**, Karinya; **Kh**, Karihona; **Kk**, Kuikuro; **Km**, Kumanakoto; **Kp**, Kaping; **Ks**, Kashuyana; **Mk**, Makushi; **Mp**, Mapoyo; **Pe**, Pemong; **Pi**, Pimenteria; **Pm**, Palmella; **Pn**, Panare; **Ti**, Tiriyo; **Tm**, Tamanaku; **Yu**, Yukpa; **Yw**, Yawarana; **Wm**, Waimiri-Atroari; **Ww**, Waiwai; **Wy**, Wayana.

Table 1 Rhythmic (iambic) stress: Tiriyo

1. Words with only light (CV) syllables, based on the stem *apoto* 'helper, servant'^a

| | | |
|---------------------------|---------------------------------|----------------------------|
| <i>apoto</i> | [(a.po):.to] | 'helper' |
| <i>m-apoto-ma</i> | [(ma.po):.to.ma] | 'you helped him' |
| <i>kit-apoto-ma</i> | [(ki.ta:).(po.to:).ma] | 'the two of us helped him' |
| <i>m-apoto-ma-ti</i> | [(ma.po:).(to.ma:).ti] | 'you all helped him' |
| <i>kit-apoto-ma-ti</i> | [(ki.ta:).(po.to:).ma.ti] | 'we all helped him' |
| <i>m-apoto-ma-po-ti</i> | [(ma.po:).(to.ma:).po.ti] | 'you all had him helped' |
| <i>kit-apoto-ma-po-ti</i> | [(ki.ta:).(po.to:).(ma.po:).ti] | 'we all had him helped' |

2. Words with at least one heavy (non-CV) syllable.

| | | |
|-------------------------|------------------------------|-----------------------------|
| <i>kin-erahta-po-ti</i> | [(ki.ne:).(rah).(ta.po:).ti] | 'he made them all be found' |
| <i>mi-repenta-ta-ne</i> | [(mi.re:).(pen).(ta.ta:).ne] | 'you all paid/rewarded him' |
| <i>m-aita-po-ta-ne</i> | [(mai).(ta.po:).ta.ne] | 'you all had it pushed' |

^aIambic feet are enclosed in parenthesis. Dots = syllable boundaries; hyphens = morpheme boundaries.

corresponds a person-marking prefix (except 1 + 3, to which correspond simple third-person markers). In some languages, the 1 + 2 prefixes were lost (Kaping, Pemong, Makushi); in others, the prefixes are replaced by pronouns as overt possessors (Yukpa, Waimiri-Atroari).

In more conservative languages, verbs have a complex inflectional system, with prefixes marking person and suffixes marking various tense-aspect-mood and number distinctions. The person-marking prefixes form what Gildea termed the *Set I* system (Table 3), variously analyzed as split-S or active-stative (e.g., by Gildea) or as cross-referencing both A (Agent) and P (Patient) (Hoff, 1968). In most languages, however, innovative systems have arisen from the reanalysis of older deverbal nominalizations or participials, and are now in competition with the *Set I* system. Most of the new systems follow ergative patterns, thus creating various cases of ergative splits and even a couple of fully ergative languages (Makushi, Kuikuro, in which the *Set I* system has been entirely lost). Gildea (1998)

Table 2 A typical Cariban pronominal system: Kashuyana

| Third person | Inanimate | | Animate | | Other persons | Sing. | Pl. |
|---------------|-------------|------------------|---------------|-----------------|---------------|---------------|----------------|
| | Sing. | Pl. | Sing. | Pl. | | | |
| Anaphoric | <i>iro</i> | <i>iro-tomu</i> | <i>noro</i> | <i>norojami</i> | 1 | <i>owi</i> | |
| Demonstrative | | | | | | | |
| Proximal | <i>soro</i> | <i>soro-tomu</i> | <i>mosoro</i> | <i>moʔtʃari</i> | 2 | <i>omoro</i> | <i>omjari</i> |
| Medial | <i>moro</i> | <i>moro-tomu</i> | <i>moki</i> | <i>mokjari</i> | 1 + 2 | <i>kumoro</i> | <i>kimjari</i> |
| Distal | <i>moni</i> | <i>mon-tomu</i> | <i>mokiro</i> | <i>mokjari</i> | 1 + 3 | <i>amna</i> | |

Table 3 Cariban person-marking systems

| Conservative (Set I) system: <i>Karinya</i> | | | | | Innovative system: <i>Makushi</i> | | | |
|---|---------------|--------------|-----------|----------------|-----------------------------------|-------|--------------|------------------|
| | 1P | 2P | 1 + 2P | 3P | (S _A) | S | P | A |
| 1A | | <i>k-</i> | | <i>s(i)-</i> | <i>Ø-</i> | 1 | <i>u(j)-</i> | <i>-u-ja</i> |
| 2A | <i>k-</i> | | | <i>m(i)-</i> | <i>m-</i> | 2 | <i>a(j)-</i> | <i>-Ø-ja</i> |
| 1 + 2A | | | | <i>kis(i)-</i> | <i>kit-</i> | 1 + 2 | <i>i-</i> | <i>-i-ja</i> |
| 3A | <i>Ø-lj-</i> | <i>a(j)-</i> | <i>k-</i> | <i>n(i)-</i> | <i>n(i)-</i> | 3Refl | <i>ti-</i> | <i>-ti(u)-ja</i> |
| (S _P) | <i>Ø-l j-</i> | <i>a(j)-</i> | <i>k-</i> | <i>n(i)-</i> | | | | |

provides a detailed account of this diachronic development.

Undersived adverbs usually take no morphology other than one nominalizing suffix. There are many postpositions, often formed with smaller locative or directional elements; they can take the same person-marking prefixes as nouns, and (usually) the same nominalizing suffix as adverbs. There are many particles in several syntactic subclasses and with various semantic and pragmatic contents (diminutives, evidentials, modals, etc.; cf. Hoff, 1986, 1990, for the *Karinya* case).

Class-changing morphology is quite rich. Verbs have many nominalizing affixes ('actual' vs. 'habitual' or 'potential' A, P, S; circumstance; action) and also adverbial-ized forms (participial, temporal, modal, etc.). There also are affixes for intransitivizing, transitivity and causativizing verb stems (according to their valency). There are several noun verbalizers (inchoative: 'to produce/have N'; privative: 'to de-N X'; dative: 'to provide X with N').

Syntax

Cariban languages are famous as examples of the rare OVS word order (Derbyshire, 1977), with *Hishkaryana* as the first case study.

- (1) *toto j-oska-je okoje* (Hishkaryana)
 man LINKER -bite-PAST snake
 'The snake bit the man.'
 (Derbyshire, 1979: 87)

Tight syntactic constituents are few: most languages have only OV-phrases (only with third-person

A and P), possessive phrases (possessor-possessed), and postpositional phrases. There are no modifier slots: 'modification' is carried out by the apposition of syntactically independent but pragmatically coreferential nominals (e.g., *the woman, that one, the tall one, the one with beads* instead of *that tall woman with beads*). Equative clauses can have a copula, but verbless clauses also occur:

- (2) *tuhu firə* (Bakairi)
 stone this
 'This is a stone.'
 (author's data)

Negation is based on a special adverbial form of the verb, derived with a negative suffix (usually *-pira*, *-pra*, *-bra*, *-ra*, etc.), in a copular clause:

- (3) *isapokara on-ene-pira aken* (Apalai)
 lizard.sp 3NEG-see-NEG 1:be:PAST
 'I did not see a *jacuraru* lizard.'
 (Lit. lizard not-seeing-it I-was)
 (Koehn and Koehn, 1986: 64)

Subordinate clauses are usually based on deverbal nominals or adverbials. In some languages, there are finite subordinate clauses (Panare, Tamanaku, Yukpa, Tiriyo). The sentences below exemplify relative clauses (in brackets): nominalizations (4) and finite clauses with relativizing particles (5).

- (4) *kaikui ə-wa:rə [pahko]* (Tiriyo)
 dog 2-known.to father
i-n-tu:ka-hpə]?
 3-PAT.NMLZR-beat-PAST
 'Do you know the dog that my father beat?'
 (author's data)

- (5) a. *tʃonkaiʔpe it-etʃeti pare* (Tamanaku)
 which 3-name priest
[n-epu-i netʃiʔ]
 3-come-PAST RELAT
 'What is the name of the priest who has (just) come?'
 (Gilij, 1782: III, 176)
- b. *akeʃ peru [kat amo=n woneta]* (Yukpa)
 that dog RELAT you=DAT 1.talk
sa=ne siuw
 thus=3.be white
 'The dog that I talked to you about was white.'
 (author's data)

With verbs of motion, a special deverbal (supine) form is used to indicate the purpose of the displacement.

- (6) *epi-be wi-tə-jai* (Wayana)
 bathe-SUPINE 1-go-PRESENT
 'I am going (somewhere) to bathe.'
 (Jackson, 1972: 60)

Lexicon and Semantics

Cariban languages have few number words, usually not specifically numerical (one = alone, lonely; two = a pair, together; three = a few); higher numbers are expressed with (often not fully conventionalized) expressions based on words for *hand*, *foot*, *person* or *body*, or are borrowings. Spatial postpositions often distinguish: vertical support ('on'), containment ('in'), attachment/adhesion, Ground properties ('in open space,' 'on summit of,' 'in water'), and complex spatial configurations ('astraddle,' 'parallel to,' 'piercing'). Some languages have 'mental state' postpositions (desiderative: *want*; cognoscitive: *know*; protective: *protective toward*; etc.). There are different verbs for eating, depending on what is eaten; to every verb corresponds a noun designating the kind of food in question (e.g., Tiriyo *ənə* 'eat meat,' *oti* 'meat food'; *enapi* 'eat fruits, vegetables', *nnapī* 'fruit, vegetable food'; *əku* 'eat bread', *uru* 'bread food'; *aku* 'eat nuts,' *mme* 'nut food').

Bibliography

- Adam L (1893). *Matériaux pour servir à l'établissement d'une grammaire comparée des dialectes de la famille caribe*. Bibliothèque Linguistique Américaine, vol. 17. Paris: J. Maisonneuve.
- Carlin E B (2004). *A grammar of Trio*. *Duisburger Arbeiten zur Sprach- und Kulturwissenschaft*, vol. 55. Frankfurt am Main: Peter Lang (Europäischer Verlag der Wissenschaften).
- Derbyshire D C (1977). 'Word order universals and the existence of OVS languages.' *Linguistic Inquiry* 8, 590–599.

- Derbyshire D C (1979). *Hixkaryana*. *Lingua Descriptive Series*, vol. 1. Amsterdam: North-Holland.
- Derbyshire D C (1985). *Hixkaryana and linguistic typology*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.
- Derbyshire D C (1999). 'Carib.' In Dixon R M W & Aikhenvald A Y (eds.) *The Amazonian languages*. Cambridge Language Surveys. Cambridge: Cambridge University Press. 23–64.
- Derbyshire D C & Pullum G K (eds.) (1986–1998). *Handbook of Amazonian languages* (4 vols). Berlin: Mouton de Gruyter.
- Durbin M (1977). 'A survey of the Cariban language family.' In Basso E (ed.) *Carib speaking Indians, culture, and society*. Tucson: University of Arizona Press.
- Gildea S (1995). 'A comparative description of syllable reduction in the Cariban language family.' *International Journal of American Linguistics* 61, 62–102.
- Gildea S (1998). *On reconstructing grammar: comparative Cariban morphosyntax*. Oxford Studies in Anthropological Linguistics, vol. 18. Oxford: Oxford University Press.
- Gilij, Filippo Salvatore (1780–1783). *Saggio di storia americana* (4 vols). Rome: Luigi Salvioni (Stampator Vaticano).
- Girard V (1971). 'Proto-Carib phonology.' Ph.D. diss., University of California, Berkeley.
- de Goeje C H (1909). *Études linguistiques caraïbes*. Verhandelingen der Koninklijke Akademie van Wetenschappen, Letterkunde, nieuwe reeks, deel X, no. 3. Amsterdam: Johannes Müller.
- de Goeje C H (1946). *Études linguistiques caraïbes*, vol. 2. Verhandelingen der Koninklijke Nederlandsche Akademie van Wetenschappen, Letterkunde, nieuwe reeks, deel II, no. 2. Amsterdam: N. V. Noord-Hollandsche Uitgeversmaatschappij.
- Greenberg J H (1987). *Language in the Americas*. Stanford: Stanford University Press.
- Hawkins R E (1998). 'Wai Wai.' In Derbyshire & Pullum (eds.). 25–224.
- Hoff B J (1968). *The Carib language*. The Hague: Martinus Nijhoff.
- Hoff B J (1986). 'Evidentiality in Carib: particles, affixes, and a variant of Wackernagel's law.' *Lingua* 69, 49–103.
- Hoff B J (1990). 'The non-modal particles of the Carib language of Surinam and their influence on constituent order.' In Payne D L (ed.) *Amazonian linguistics: studies in lowland South American languages*. Austin: University of Texas Press. 495–541.
- Jackson W S (1972). 'A Wayana grammar.' In Grimes J E (ed.) *Languages of the Guianas*. Norman: Summer Institute of Linguistics and University of Oklahoma Press. 47–77.
- Kaufman T K (1994). 'The native languages of South America.' In Moseley C & Asher R E (eds.) *Atlas of the world's languages*. New York: Routledge. 46–76.
- Koehn E & Koehn S (1986). 'Apalai.' In Derbyshire & Pullum (eds.). 33–127.
- Loukotka Č (1968). *Classification of South American Indian languages*. Los Angeles: Latin American Center, University of California.

- Meira S (1998). Rhythmic stress in Tiriyó. *International Journal of American Linguistics* 64, 352–378.
- Meira S (1999). ‘Syllable reduction and ghost syllables in Tiriyó.’ In Hwang S J & Lommel A R (eds.) *XXV LACUS Forum*. Fullerton, CA: The Linguistic Association of Canada and the United States (LACUS). 125–131.
- Meira S (2000). *A reconstruction of Proto-Taranoan: phonology and morphology*. Munich: LINCOM Europa.
- Meira S (2005). *A grammar of Tiriyó*. Berlin: Mouton de Gruyter.
- Meira S & Franchetto B (2005). ‘The southern Cariban languages and the Cariban family.’ *International Journal of American Linguistics*. 71, 127–192.
- Muller M C M (1994). *Diccionario ilustrado panare-español, español-panare*. Caracas: Comisión Quinto Centenario, Gráficas Armitano.
- Rodrigues A D (1985). ‘Evidence for Tupi-Cariban relationship.’ In Klein H & Stark L (eds.) *South American languages: retrospect and prospect*. Austin: University of Texas Press. 371–404.
- Shafer R (1963). *Vergleichende Phonetik der karaibischen Sprachen*. Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Letterkunde, nieuwe reeks, deel LXIX, no. 2. Amsterdam: N. V. Noord-Hollandsche Uitgeversmaatschappij.

Catalan

M W Wheeler, University of Sussex, Brighton, UK

© 2006 Elsevier Ltd. All rights reserved.

Geography and Demography

The territories where Catalan is natively spoken cover 68 730 km², of which 93% lies within Spain (see Figure 1). They are:

1. The Principality of Andorra
2. In France: North Catalonia – almost all of the département of Pyrénées-Orientales
3. In Spain: Catalonia, except for the Gascon-speaking Vall d’Aran; the eastern fringe of Aragon; most of Valencia (the Comunitat Valenciana), excepting some regions in the west and south that have been Aragonese/Spanish-speaking since at least the 18th century; El Carxe, a small area of the province of Murcia, settled in the 19th century; and the Balearic Islands
4. In Italy: the port of Alghero (Catalan *L’Alguer*) in Sardinia

Table 1 shows the population of these territories (those over 2 years of age in Spain) and the percentages of the inhabitants who can understand, speak, and write Catalan. Information is derived from the 2001 census in Spain together with surveys and other estimates; the latter are the only sources of language data in France and Italy. The total number of speakers of Catalan is a little under 7.5 million. Partly as a result of the incorporation of Catalan locally into the education system, there are within Spain a significant number of second-language speakers who are included in this total. Virtually all speakers of Catalan are

bilingual, using also the major language of the state they live in. (Andorrans are bilingual in Spanish or French, or are trilingual.)

Genetic Relationship and Typological Features

Catalan is a member of the Romance family and a fairly prototypical one, as befits its geographically central position in the European Romance area. Some particularly noteworthy characteristics are pointed out here (for more details see Wheeler, 1988). In historical phonology, note the palatalization of initial /l-/ and loss of stem-final /n/ that became word final, for example, LEONEM > *lleó* [lɛˈo] ‘lion.’ Original intervocalic -C-, -TJ-, -D- became /w/ in word-final position and were lost elsewhere, for examples, PLACET > *plau* [ˈplaw] ‘please.3.SING,’ PLACEMUS > *plaem* [pləˈem] ‘please.1.PL.’ As the previous examples also illustrate, post-tonic nonlow vowels were lost, so that a dominant pattern of phonological words is of consonant-final oxytones. The full range of common Romance verbal inflection is retained, including inflected future (*sentirà* ‘hear.3.SING.FUT’), widely used subjunctives, and a contrast between present perfect (*ha sentit* ‘has heard’) and past perfective (*sentí* ‘heard.3.SING.PERF’). In addition to the inherited past perfective form, now largely literary, Catalan developed a periphrastic past perfective using an auxiliary that was originally the present of ‘go’ (*va sentir* ‘AUX.PERF.3.SING hear.INF’). In some varieties of Catalan, this construction has developed a subjunctive (*vagi sentir* ‘AUX.PERF.SUBJ.3.SING hear.INF’), introducing, uniquely in Romance, a perfective/imperfective aspect distinction in the subjunctive. Considerable use

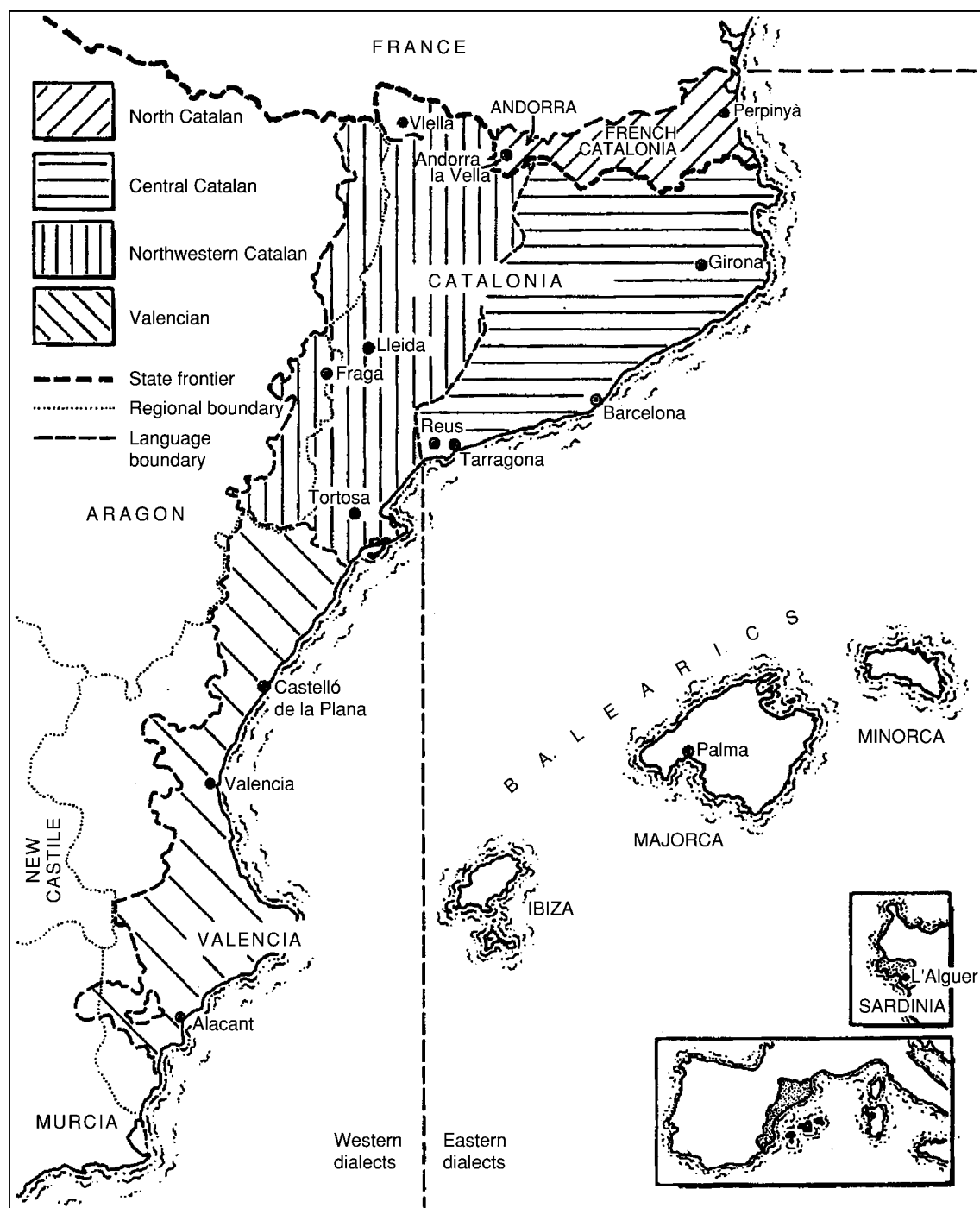


Figure 1 Catalan-speaking areas and dialects.

is made of pronominal and adverbial clitics that attach to verb forms in direct and indirect object functions or partitive or adverbial functions, quite often in clusters of two or three, as in (1).

- (1) us n'hi envi-en
 2.PL.OBJ PART.LOC send-3.PL
 "they send some to you (PL)
 there"

Most of the pronominal/adverbial clitics have several contextually conditioned forms; thus, the partitive clitic shows variants *en* ~ *n'* ~ *-ne*. Clitic climbing is commonly found with a pronominal complement of a verb that is itself the complement of a (semantic) modal, as in (2). This example also shows the (optional) gender agreement of a perfect participle with a preceding direct object clitic.

Table 1 Catalan language demography and competences

| <i>Territory</i> | <i>Population</i> | <i>Understand Catalan (%)</i> | <i>Speak Catalan (%)</i> | <i>Write Catalan (%)</i> |
|------------------|-------------------|-----------------------------------|------------------------------|------------------------------|
| Andorra | 66 000 | 97 | 91 | (No data) |
| North Catalonia | 363 000 | 59 | 41 | 10 |
| Catalonia | 6 215 000 | 95 | 75 | 50 |
| Aragon fringe | 50 000 | 95 | 90 | (No data) |
| Valencia | 4 145 000 | 85 | 48 | 23 |
| Balearics | 822 000 | 90 | 68 | 26 |
| Alghero/L'Alguer | 38 000 | 53 | 46 | (No data) |
| Total | 11 699 000 | 89 | 64 | 37 |

(2) no l'he sab-ud-a agafa-r
 not DO.3.SING.F know- catch-INF
 have.1.SING PART-F
 "I haven't been able to catch it (FEM)"

A fair number of items in the basic vocabulary are etymologically distinct from the corresponding terms in neighboring Romance languages, for example, *estimar* 'to love,' *ganivet* 'knife,' *gens* 'not at all,' *massa* 'too,' *pujar* 'to go up,' *tardor* 'autumn,' and *tou* 'soft.'

Dialects

Although there are significant dialect differences in Catalan, the dialects are to a high degree mutually intelligible. They are conventionally divided into two groups, on the basis of differences in phonology as well as some significant features of verb morphology; there are some interesting lexical differences, too. The eastern dialect group (see **Figure 1**) includes North Catalan or *rossellonès* (in France), central Catalan (in the eastern part of Catalonia), Balearic, and *alguerès* (in Alghero/L'Alguer). The western group consists of Northwestern Catalan (western and southern Catalonia and eastern Aragon) and Valencian. The main diagnostic heterogloss distinguishing the two major dialect groups involves vowel reduction in unstressed syllables: In the eastern dialects /a/ is pronounced [ə] in unstressed syllables and, with some exceptions, /e/ and /ɛ/ are also reduced to [ə], whereas /o/ and /ɔ/ are reduced to [u].

History

Catalan is a variety of Latin that developed originally on a small territory on either side of the eastern Pyrenees. Expansion of this territory, the Marca Hispanica of the Carolingian empire, is associated with a process of developing political independence, beginning with the separation (A.D. 988) of the county of

Barcelona from the trunk of the Carolingian domain. Eventual fusion with the crown of Aragon (1162) gave new momentum to this projection. In 1151, a treaty between the kings of Aragon and Castile had carved up the future conquest of territories then under Arab control, so that Valencia would fall to the crown of Aragon while lands further west would be attached to Castile. The kingdom of Valencia was captured in the 1230s and was populated by speakers from various parts of Catalonia and Aragon, although a numerous subordinate population of Arabic-speaking *moriscos*, as they were called, remained until their expulsion in 1609. The Balearic Islands were conquered between 1229 and 1287 and were resettled by speakers largely from eastern Catalonia. Sicily was also captured for the house of Barcelona (1282), as was Sardinia (1323–1327); Catalan was widely used as an official language in Sicily until the 15th century and in Sardinia until the 17th century. In Sardinia, only the port of Alghero was subject to Catalan resettlement, and it has remained Catalan-speaking to the present day. The original expansion southward of Catalan following the reconquest extended as far as Murcia and Cartagena, although the kingdom of Murcia became Spanish-speaking during the 15th century.

The chancellery of the kingdom of Aragon was trilingual, using Latin, Catalan, and Aragonese as the occasion required. A substantial body of Catalan literature in various prose and verse genres was produced before decline set in in the 16th century. In 15th-century Valencia the court was already bilingual, and after the merger of the Aragonese and Castilian crowns in 1479 Spanish (Castilian) gradually increased in prestige throughout the Catalan territories, with the urban and literate classes becoming bilingual. From the 16th century, Catalan came increasingly under Spanish influence in vocabulary, syntax, pronunciation, and orthography as a result of the social and cultural prestige of Castile. It was not until the 19th century that a substantial Catalan literary and cultural revival took place,

which continues to the present. Standardization of the modern language was achieved in the early 20th century.

Since the Second World War, most of the Catalan-speaking territories have experienced a substantial immigration of non-Catalan speakers. In France, these have been *pieds noirs* resettled from Algeria and retired people from various parts of France. In Catalonia and Valencia, the population almost doubled between 1950 and 1975 as people from less-developed southern Spain sought employment in the manufacturing and service industries. Majorca and Ibiza (Eivissa) have attracted a workforce from many parts of Spain, feeding the tourist industry. Many immigrants have wished to acquire Catalan, or at least have wished their children to do so, as an aid to integration, but until the late 1970s there were few opportunities to realize this. These large Spanish-speaking communities have added to the institutional and cultural pressures in favor of the use of Spanish in the Catalan territories.

In 1659, Philip IV of Spain ceded the northern part of Catalonia (essentially the modern département of Pyrénées-Orientales) to the French crown. From that point, North Catalonia became subject to the linguistic unification policies of the French state. French became the official language in 1700 and has had a marked influence on the vocabulary of North Catalan and, in recent times, on its phonology as well. Minorca was under British rule during most of the 18th century, and there is a handful of Minorcan Anglicisms in the vocabulary dating from that period. The dialect of Alghero is, not surprisingly, heavily influenced by Sardinian and even more so by Italian in all components of the language.

Present Sociolinguistic Situation

The status, situation, and prospects of the Catalan language are significantly different in each of the territories in which it is spoken, although each of those in Spain shares, in some way, the consequences of Catalan's having been for centuries an oppressed minority language. The cultural decline and loss of prestige affecting Catalan from the 16th century onward has already been mentioned. The defeat of the Catalans in the war of the Spanish Succession (1714) initiated a series of measures, extending throughout the 18th and 19th centuries, that imposed the use of Spanish in public life, for example, in accounts, in preaching, in the theater, in the criminal courts, in education, in legal documents, in the civil registers, and on the telephone. In the 20th century, these measures were mostly repeated and supplemented by the imposition of Spanish in catechism, by the

prohibition of the teaching of Catalan, and by sanctions against people refusing to use Spanish. The Second Republic (1931–1939) to a large extent removed these restrictions, but Franco's victory in the Spanish Civil War was followed in 1940 by a total ban on the public use of Catalan. Despite a gradual relaxation allowing some publication of books and magazines, Catalan remained excluded from nearly all public institutions until Spain's adoption of a democratic constitution in 1978.

In the early 1980s, Catalonia, Valencia, and the Balearics obtained their statutes of autonomy, involving co-official status for Spanish and Catalan. All of these statutes promote language normalization, the goal of which is universal bilingualism without diglossia. In Catalonia, the expressed aim of the Generalitat (the autonomous government) goes further than this: It seeks to make the local language the normal medium of public life, with Spanish having a secondary role as an auxiliary language or a home language for its native speakers. In Catalonia, the teaching of Catalan is obligatory in all schools, and primary and secondary education through the medium of Catalan now reaches at least 60% of the population. In Valencia and the Balearics, the *de facto* policy has been to promote effective knowledge of Catalan through education and to enhance its status while largely preserving a diglossic relationship between Spanish and Catalan. In Valencia, significant political forces reject the name Catalan for the local language and insist on the term Valencian. Although the Balearic Islands Council passed a linguistic normalization law in 1986, progress has been inconsistent, although Catalan is widely available in the education system which includes some Catalan-medium education.

In Andorra, Catalan has always been the sole official language. In 1993, Andorra adopted a new constitution, and the government has been pursuing an active Andorranization policy, involving Catalan-medium education. The status of Catalan in North Catalonia is parallel to that of the other traditional minority languages in France. Language shift was all but universal after the Second World War, so that most native speakers are (as of 2004) over 60 years old. Catalan has at best an occasional, decorative role in public life. In primary schools, some 30% study Catalan (as a foreign language) and, in secondary schools, some 15%.

The current trend is for intergenerational language shift from Catalan in French Catalonia, in Alghero, in southern Valencia around Alicante (Alacant), and possibly in Palma (Majorca). Elsewhere, Catalan is holding its own, with some evidence of intergenerational shift toward Catalan in Catalonia.

Bibliography

- Badia i Margarit A M (1951). *Gramàtica històrica catalana*. Barcelona, Spain: Noguer. [Catalan translation *Gramàtica històrica catalana*. Valencia: 3 i 4, 1981.]
- Moll F de B (1952). *Gramàtica històrica catalana*. Madrid: Gredos. [Catalan translation *Gramàtica històrica catalana*. Valencia: Universitat, 1991.]
- Nadal J M & Prats M (1982–1996). *Història de la llengua catalana* (2 vols.). Barcelona, Spain: Edicions 62.
- Pradilla M À (ed.) (1999). *La llengua catalana al tombant del mil·lenni: aproximació sociolingüística*. Barcelona, Spain: Empúries.

- Solà J, Lloret M R, Mascaró J & Pérez Saldanya M (eds.) (2002). *Gramàtica del català contemporani* (3 vols.). Barcelona, Spain: Empúries.
- Wheeler M W (1988). 'Catalan.' In Harris M & Vincent N (eds.) *The Romance languages*. London: Routledge. 170–208.
- Wheeler M W (2005). *The phonology of Catalan*. Oxford: Oxford University Press.
- Wheeler M W, Yates A & Dols N (1999). *Catalan: a comprehensive grammar*. London: Routledge.

Caucasian Languages

B G Hewitt, SOAS, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Around 38 languages are deemed to be indigenous to the Caucasus; often difficult demarcation between language and dialect explains the uncertainty. The ancestral homelands are currently divided between:

1. Russia's north Caucasian provinces (Circassian, Abaza, Ingush, Chechen, Avaro-Ando-Tsezic, Lako-Dargic, northern Lezgic);
2. *de facto* independent Abkhazia (Abkhaz, Mingrelian, Svan, Georgian, Laz);
3. Georgia (Georgian, Mingrelian, Svan, Laz, Bats, Chechen, Avar, Udi);
4. Azerbaijan (Lezgi, Budukh, Kryts', Khinalugh, Rutul, Ts'akhur, Avar, Udi) Turkey (Laz, Georgian).

Diaspora-communities of North (especially north-west) Caucasians can be found across former Ottoman territories, particularly Turkey, where the majority Circassian and Abkhazian populations reside and where the term 'Cherkess' often indiscriminately applies to any North Caucasian. Circassians are found in Syria, Israel, and Jordan, home also to a significant Chechen population. Speaker numbers range from 500 (Hinukh) to 3–4 million (Georgian). Many of the languages are endangered.

Three families are usually recognized:

- A. South Caucasian (Kartvelian)
- Georgian
 - Svan
 - Mingrelian (Megrelian)
 - Laz (Ch'an)

[Scholars in Georgia regard Mingrelian and Laz as codialects of Zan]

- B. North West Caucasian

Abkhaz

Abaza

Ubykh (extinct from 1992)

West Circassian (Adyghe)

East Circassian (Kabardian)

- C. Nakh-Daghestanian

- (a) *Nakh (North Central Caucasian)*

Chechen

Ingush

Bats (Ts'ova Tush)

- (b) *Daghestanian (North East Caucasian)*

1. Avaro-Ando-Tsezic(/Didoic):

Avaric: **Avar**

Andic: Andi, Botlikh, Godoberi, K'arat'a (Karata), Akhvakh, Bagvalal, T'indi (Tindi), Ch'amalal (Chamalal)

Tsezic: Tsez (Dido), Khvarshi, Hinukh, Bezht'a (Bezhta) (K'ap'uch'a), Hunzib (these last two are sometimes regarded as codialects)

2. Lako-Dargic:

Lakic: **Lak**

Dargic: **Dargwa** (Dargi(n)) – some treat K'ubachi, Chiragh, and Megeb as full languages

3. Lezgic:

Lezgi(an), **Tabasaran** (Tabassaran), Rutul (Mukhad), Ts'akhur (Tsakhur), Aghul, Udi, Archi, Budukh, Khinalugh, Kryts' (Kryts)

Some challenge the Lezgic status of Archi, Khinalugh, Budukh, and Kryts.' Mutual intelligibility basically

exists between Laz and Mingrelian, Abkhaz, and Abaza, West and East Circassian. Only Georgian has an ancient tradition of writing, but during the Soviet period the languages in bold all enjoyed literary status. Publishing in Mingrelian, Laz, Ts'akhur, Aghul, Rutul, and Udi was tried in the 1930s but discontinued, though there have been some post-Soviet attempts to publish more widely (including Dido).

Phonetics and Phonology

All Caucasian languages have voiced vs. voiceless aspirate vs. voiceless ejective plosives, affricates, and occasionally fricatives, to which some add a *fortis* series (voiceless unaspirated or geminate). North West Caucasian is characterized by large consonantal inventories coupled with minimal vowel systems, consisting of at least the vertical opposition open /a/ vs. closed /ə/. Ubykh possessed 80 phonemes (83 if the plain velar plosives attested only in loans are admitted), with every point of articulation between lips and larynx utilized and displaying the secondary features of palatalization, labialization, and pharyngalization – Daghestanian pharyngalization is normally assigned to vowels (Table 1).

Some recent analyses of Daghestanian languages have produced inventories rivaling those of the North West Caucasian, though no parallel minimality among the vowels is posited. One analysis of Archi assigns it 70 consonants (Table 2).

Table 1 Consonantal phonemes for Ubykh

| | | | | | |
|-----------------|-----------------|------------------|-----------------|-----------------|----------------|
| p | b | p' | | m | w |
| p ^ʕ | b ^ʕ | p ^ʕ | | m ^ʕ | w ^ʕ |
| | | | f | | |
| | | | v ^ʕ | | |
| t | d | t' | | n | r |
| t ^w | d ^w | t' ^w | | | |
| ts | dz | ts' | s | z | |
| ts ^w | dz ^w | ts' ^w | ɕ | ʒ | |
| tʃ | dʒ | tʃ' | ɕ ^w | ʒ ^w | |
| tʃ ^w | dʒ ^w | tʃ' ^w | ʃ | ʒ | |
| | | | ʃ ^w | ʒ ^w | |
| ts̥ | dz̥ | ts̥' | ʂ | ʐ | |
| | | | ʂ ^w | ʐ ^w | |
| | | | ʂ ^ʕ | ʐ ^ʕ | |
| (k) | (g) | (k') | x | ɣ | |
| k ^ʕ | g ^ʕ | k' ^ʕ | | | |
| k ^w | g ^w | k' ^w | | | |
| q | | q' | χ | ʁ | |
| q ^ʕ | | q' ^ʕ | χ ^ʕ | ʁ ^ʕ | |
| q ^j | | q' ^j | χ ^j | ʁ ^j | |
| q ^w | | q' ^w | χ ^w | ʁ ^w | |
| q ^{ʕw} | | q' ^{ʕw} | χ ^{ʕw} | ʁ ^{ʕw} | |
| | | | h | | |

Noticeable here, is the presence of 10 laterals, though some specialists recognize no more than three or four.

Kartvelian occupies a mid-position with between 28 and 30 consonants (see Georgian). Georgian shares with Avar and Andi the simple five-vowel triangle (Table 3).

Schwa is added to this in the other Kartvelian languages, while the various Svan dialects have length and/or umlaut, Upper Bal having the richest system (Table 4).

Triangular or quadrilateral vowel systems are attested in Nakh-Daghestanian (Table 5).

All but /y, ε, œ/ possess long counterparts, and the nasalized vowels: /ĩ, ẽ, ã, ẽ̃, û, û̃, ã̃, ỹ̃/ have also been recognized. Table 6 shows the Hunzib basic vowels.

Table 2 Consonantal system of Archi

| | | | | | | | |
|-----------------|----------------|-------------------|-----------------|----------------|-----------------|----------------|---|
| p | b | p' | p: | | | m | w |
| t | d | t' | t: | | | n | r |
| t ^w | d ^w | | | | | | |
| ts̥ | | ts̥' | ts: | s | s: | z | |
| ts ^w | | ts ^w ' | | s ^w | s: ^w | z ^w | |
| tʃ̥ | | tʃ̥' | tʃ: | ʃ | ʃ: | ʒ | |
| tʃ ^w | | tʃ ^w ' | | ʃ ^w | ʃ: ^w | ʒ ^w | |
| k̥ | | k̥' | | ɸ | ɸ: | ɸ̥ | l |
| k ^w | | k ^w ' | | ɸ ^w | ɸ: ^w | | |
| | | | | | | | j |
| k | g | k' | k: | | | | |
| k ^w | g ^w | k ^w ' | k: ^w | | | | |
| q | | q' | q: | χ | χ: | ʁ | |
| q ^w | | q ^w ' | | χ ^w | χ: ^w | ʁ ^w | |
| | | | | h | | ɸ̥ | |
| | | ʔ | | | | | |

Table 3 Georgian-Avar-Andi vowel system

| | | | | |
|---|---|---|---|---|
| i | | | | u |
| | ε | | ɔ | |
| | | a | | |

Table 4 Svan's upper Bal vowel system

| | | | | | | |
|---|----|---|----|---|----|----|
| i | i: | y | y: | | u | u: |
| | | | | ə | ə: | |
| ε | ε: | œ | œ: | | ɔ | ɔ: |
| a | a: | | | | ɑ | ɑ: |

Table 5 Bezht'a basic vowel system

| | | | | |
|---|---|---|---|---|
| i | y | | | u |
| | ε | œ | | ɔ |
| | | a | ɑ | |

All these Hunzib vowels have long counterparts, and fluctuating nasalization on short vowels has been observed.

The simplest (near-)quadrilateral system is attested in Chiragh Dargwa, with four pairs distinguished by length (Table 7). Udi has been analyzed in Table 8, whilst Chechen presents the complicated system (Table 9).

Most, if not all, of these can be nasalized as a result of the weakening of a following /n/.

Stress is sometimes distinctive (Abkhaz-Abaza) but usually not. Tonal distinctions have been proposed for some of the Daghestanian languages (Andi, Akhvakh, Ch'amalal, Khvarshi, Hinukh, Bezht'a, Tabasaran, Ts'akhur, Ingush, and Budukh).

Table 6 Hunzib basic vowel system

| | | | | | |
|---|---|---|---|---|---|
| i | | | | ɨ | u |
| | ɛ | | ə | ɔ | |
| | | ɑ | ɒ | | |

Table 7 Chiragh Dargwa vowel system

| | |
|------|------|
| i(:) | u(:) |
| ɛ(:) | ɑ(:) |

Table 8 Udi vowel system

| | |
|----------------------|------------------|
| i i ^ʰ (y) | u u ^ʰ |
| ɛ ɛ ^ʰ (œ) | ɔ ɔ ^ʰ |
| (a) | ɑ ɑ ^ʰ |

Table 9 Chechen vowel system

| | |
|-------------|-------|
| i i: y y: | u u: |
| je ie qœ yœ | wo uo |
| e e: | o o: |
| a a: | ɑ ɑ: |

Table 10 Avar locative case endings

| Series | Essive | Allative | Ablative |
|-----------------|---------------------|----------|------------|
| 1. 'on' | -d(.)ɑ | -d.ɛ | -d(.)ɑ.s:ɑ |
| 2. 'near' | -q: | -q:.ɛ | -q:.ɑ |
| 3. 'under' | -kʰ: | -kʰ:.ɛ | -kʰ:.ɑ |
| 4. 'in (mass)' | -kʰ: | -kʰ:.ɛ | -kʰ:.ɑ |
| 5. 'in (space)' | -D (= class-marker) | -D-ɛ | -s:ɑ |

Morphology

North West Caucasian sememes are typically C(C)(V), and minimal case systems combine with highly polysynthetic verbs, which may contain up to four agreement prefixes, locational preverbs, orientational preverbs and/or suffixes, interrogative and conjunctive elements, and markers of tense-modality, (non-)finiteness, causation, potentiality, involuntariness, polarity, reflexivity, and reciprocity (see Abkhaz). Kartvelian balances a moderate total of cases with reasonably complex verbs, which may contain: agreement with two or three (rarely four) arguments via two sets of agreement affixes, directional/perfectivizing preverbs (the large total in Mingrelian-Laz suggests North West Caucasian influence), and markers of tense-aspect-modality, causation, potentiality, version (vocalic prefixes indicating certain relations between arguments), and voice – Kartvelian is the only family to have a full active-passive diathetic opposition. Nakh-Daghestanian has complex nominal systems with both grammatical and sometimes large numbers of locative cases; Lezgi(an), Aghul, and Udi apart, nouns fall into one of between two and (depending on the analysis) five or eight (largely covert) classes. Verbs are correspondingly simple: agreement is totally absent from Lezgi(an) and Aghul; elsewhere, verbs with an agreement slot typically allow only class agreement (Andic), though some languages (Bats, Lak-Dargwa, Tabasaran, Akhvakh, Archi, Hunzib, and Avar dialects) have added perhaps rudimentary person agreement, whilst Udi has person agreement only. Some languages have a small selection of preverbs. Some distinguish perfective from imperfective roots. Some North Caucasian verbs can be construed transitively or intransitively (?passively), depending on the clausal structure. Antipassives are also attested.

Avar illustrates a typical system of locative-cases (Table 10).

Ergativity and some other oblique case function are often merged in a single morph.

Deictic systems range from two-term (Mingrelian, Ubykh, Kryts'), through three-term (Georgian, Abkhaz, Circassian), to five-term in a swathe of Daghestanian, and even six-term (Lezgi(an), Godoberi).

Counting systems are predominantly vigesimal, at least up to '99' (though Bats is vigesimal throughout), but some systems are decimal.

Syntax

Word orders are: Kartvelian and Nakh-Daghestanian AN, GN, N-Postposition, SOV, though Old Georgian was rather NA and NG; North West Caucasian GN, predominantly NA, N-Postposition, SOV. Some degree of ergativity characterizes all the languages, but in Mingrelian, where the system was originally as illustrated for Georgian (*q.v.*), the ergative case marker was extended vertically to replace the original nominative for intransitive (including indirect) verbs in Series II (aorist indicative and subjunctive), where it functions as a Series II nominative allomorph, the original nominative effectively becoming an accusative just for Series II. Laz has extended the case marker horizontally across its three series for all transitive subjects. Active-inactive alignment plays a role in some languages (Bats).

A nominative/absolutive argument is the obligatory minimum in a clause, and where verbs have class agreement, this is the determiner for the class marker (which in some languages also appears on adverbs and as part of a locative case exponent); the determiner for person agreement in languages with class agreement might be this same or a different argument (e.g., the logical subject), depending on a variety of factors.

Verbs such as *want*, *have*, *hear* are construed indirectly with the logical subject in an oblique case, but, if Kartvelian and North West Caucasian employ just the dative/general oblique case for this argument, greater distinctions can apply in Nakh-Daghestanian: Avar employs its dative case with verbs of emotion (*love*), a locative (Series I essive) with verbs of perception (*see*), and the genitive for the possessor in conjunction with the copula.

Only Kartvelian has the category of subordinating conjunctions, naturally associated with full clauses containing indicative or subjunctive finite verbs. Such structures are rare in North Caucasian, where one finds a variety of nonfinite (nominalized) verb forms fulfilling the subordinate role.

Examples:

ilu-di rikʰi'i b-ɛ3-a vs. rikʰi'i b-ɛ3-a
mother- meat. 3-fry-
Erg Absol₃ Past
 'Mother fried the meat' vs. 'The meat (was) fried'
 (Andi)

is-t'i sii kart:ɔl-tʰa
brother-Erg water.Absol boil-Pres
 'Brother is boiling the water' (Bezht'a)

vs.

is sii-d kart:ɔl-da-tʰ
brother.Absol water-Instr boil-AntiPass-Pres
 'Brother is regularly engaged in boiling water'
 (Bezht'a)

k'ɔf-k dʒkab-i kɔ-ø-dʒir-u
man-Nom_A girl-Acc_B Prev-her_B-see-he.Aor_A

vs.

dʒkab-k dɔ-kur-u
girl-Nom_A Prev-die-she.Aor_A
 'The man saw the girl' vs. 'The girl died'
 (Mingrelian)

k'ɔf-s dʒkab-i ø-a-dʒir-ε
man-Dat_B girl-Nom_A he_B-Pot-see-her.Pres_A
 'The man can see the girl' (Mingrelian)

vs.

k'ɔf-s dʒkab-k k-ø-a-dʒir-u
man-Dat_B girl-Nom_A Prev-he_B-Pot-see-her.Aor_A
 'The man could see the girl' (Mingrelian)

ins:u-je j.as j-ɔkʰi'-u-la
father-Obl-Dat daughter₂.Absol 2-love-TV-Pres
 'Father loves (his) daughter' (Avar)

ins:u-d.a w.as-ul r-ix:u-la
father-Obl-LocI son-Pl.Absol Pl-see-TV-Pres
 'Father sees (his) sons' (Avar)

ins:u-l tʃu b-ugɔ
father-Obl-Gen horse₃.Absol 3-be.Pres
 'Father has a horse' (Avar)

lamʃged-χen-iʃ bikw-d sga
shade-from-Gen wind-Erg_A Prev
 la-ø-j-k'wif-ø, εre
Prev-it_B-SV-admit-it.Aor_A that

mine uʃχwar nensga
their each.other.Dat between
 χ.ɔ-l.qmaʃ-a miʒ
CompPref-strong-CompSuff sun.Nom_A

lɔ.m.ar-ø
apparently.be-it_A
 'The north wind admitted that the sun was
 apparently the stronger of them'
 (Lower Bal Svan)

təka-ʒə-m təka-r jaʒ naħ.rəj naħ
sun-wind-the. sun-the. self much more
Erg/Obl_{III} Absol_I

ø-za.rə-ʒaʃə-r ø-qə-gʷə.rə-ø-mə-ɾʷa-ma
it_I-how-strong- it_I-Prev-Prev-it_{III}-not-
Absol.N/E. admit.
Stat.Pres_I N/F-if_I

ø-mə-χʷə-n-aw ø-χʷə-ka
it_I-not-happen-Fut-Abs_I it_I-happen-Aor.Fin
 'It became impossible for the north wind not to
 admit how/that the sun is stronger than it'
 (Temirgoi West Circassian)

Kinship

Kartvelian is unrelated to any known language or language family, but the debate continues concerning the relationship between the northern families. Linkage to Hattic is postulated for northwestern Caucasian and to Hurrian for Nakh-Daghestanian. Udi has recently been conclusively demonstrated to descend from Caucasian Albanian.

Bibliography

- Berg H van den (1995). *A grammar of Hunzib (with texts and lexicon)*. Lincom Studies in Caucasian Linguistics 01. Munich: Lincom Europa.
- Berg H van den (2001). *Dargi folktales. Oral stories from the Caucasus and an introduction to Dargi grammar*. Leiden: Research School CNWS.
- Berg H van den (2004). 'The East Caucasian language family.' Special Edition of *Lingua*. 147–190.
- Boeder W (1979). 'Ergative syntax and morphology in language change: the South Caucasian languages.' In Plank F (ed.) *Ergativity*. New York: Academic Press. 435–480.
- Boeder W (2004). 'South Caucasian.' Special edition of *Lingua*. 5–89.
- Catford J C (1976). 'Ergativity in Caucasian languages.' In *Papers from the 6th Meeting of the Northeast Linguistics Society*. Montreal. NELS, 6, 37–48.
- Catford J C (1977). 'Mountain of tongues: the languages of the Caucasus.' *Annual Review of Anthropology* 6, 283–314.
- Charachidzé G (1981). *Grammaire de la langue Avar*. Paris: Editions Jean-Favard.
- Chirikba V A (1996). *Common West Caucasian. The reconstruction of its phonological system and parts of its lexicon and morphology*. Leiden: CNWS.
- Deeters G (1930). *Das Kharthwelische Verbum*. Leipzig: Kommissionsverlag von Markert und Petters.
- Dumézil G (1975). *Le verbe Oubykh. Etudes descriptives et comparatives (avec la collaboration de Tefvik Esenç)*. Paris: Klincksieck.
- Dumézil G (1976). 'Compléments et corrections au Verbe Oubykh (1975) I.' *Bedi Kartlisa (revue de kartvélogie)* XXXIV, 9–15.
- Greppin J (ed.) (1989–2004). *The indigenous languages of the Caucasus*. 1: *Kartvelian languages* (Harris A C [ed.], 1991; 2: *North West Caucasus* (Hewitt B G [ed.], 1989; 3: *North East Caucasian languages, Part 1* (Job D M [ed.], 2004; 4: *North East Caucasian languages, Part 2* (Smeets R [ed.], 1994. Delmar, New York: Caravan Books.
- Harris A C (2002). *Endoclititics and the origins of Udi morphosyntax*. Oxford: OUP.
- Haspelmath M (1993). *A grammar of Lezgian*. Berlin: Mouton de Gruyter.
- Hewitt B G (1987). *The typology of Subordination in Georgian and Abkhaz*. Berlin: Mouton De Gruyter.
- Hewitt B G (ed.) (1992). *Caucasian perspectives*. Munich: Lincom Europa.
- Hewitt B G (1998). 'Caucasian languages.' In Price G (ed.) *Encyclopaedia of the languages of Europe*. 57–81. [Paperback version 2000.]
- Hewitt B G (2004). 'North West Caucasian.' Special Caucasian edition of *Lingua*. 91–145.
- Hewitt B G (2004). *Introduction to the study of the Caucasian languages*. Munich: Lincom Europa.
- Kibrik A E & Kodzasov S V (1988). *Sopostavitel'noe izuchenie dagestanskix jazykov. Glagol* [Contrastive study of the Daghestanian languages. The verb]. Moscow: University Press.
- Kibrik A E & Kodzasov S V (1990). *Sopostavitel'noe izuchenie dagestanskix jazykov. Imja. Fonetika* [Contrastive study of the Daghestanian languages. The noun. phonetics]. Moscow: University Press.
- Klimov G A (1994). *Einführung in die kaukasische Sprachwissenschaft*, translated and expanded by Gippert, J. Hamburg: Buske.
- Klimov G A (1998). *Trends in linguistics. Documentation 16. Etymological dictionary of the Kartvelian languages*. Berlin: Mouton De Gruyter.
- Kuipers A H (1960). *Phoneme and morpheme in Kabardian*. 'S-Gravenhage: Mouton.
- Nichols J (1997). 'Chechen phonology.' In Kaye A S (ed.) *Phonologies of Asia and Africa (including the Caucasus)*, vol. 2. Winona Lake: Eisenbrauns. 941–971.
- Nikolayev S L & Starostin S A (1994). *A North Caucasian etymological dictionary*. Moscow: Asterisk.
- Paris C (1974). *Système phonologique et phénomènes phonétiques dans le parler besney de Zennun Köyü (Tcherkesse oriental)*. Paris: Klincksieck.
- Smeets R (1984). *Studies in West Circassian phonology and morphology*. Leiden: Hakuchi Press.
- Tuite K (1998). *Kartvelian morphosyntax: number agreement and morphosyntactic orientation in South Caucasian languages*. Studies in Caucasian Linguistics 12. Munich: Lincom Europa.

Cebuano

J U Wolff, Cornell University, Ithaca, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Cebuano is spoken in the central and southern Philippines. It is a member of the Austronesian family of languages, the group of languages spoken throughout most of Indonesia, northward into the Philippines and Taiwan and eastward through much of Papua New Guinea and over the Pacific as far as Hawaii and Easter Island. The languages of the Philippines, with the exceptions of the Spanish Creoles, Chabacano and Chavacano, are closely related and typologically similar to one another. In particular, Cebuano is subgrouped with Tagalog and is similar to Tagalog in much the same way as Italian and Spanish are similar to each other (see **Tagalog**). Cebuano is called *Sinugbaʔanun* or *Sinibuanu* natively, and is sometimes referred to as 'Sugbuanon' in the literature about the language. Cebuano is also commonly called 'Visayan' (*Binisayaʔ* natively), after the name of the region of the central Philippines. However, there are in fact more than 30 languages spoken in this area, all of which are referred to as 'Visayan,' such that many publications referring to 'Visayan' have to do with languages other than Cebuano.

Cebuano is spoken by somewhere around a fifth of the population of the Philippines. It is thus second only to Tagalog in number of speakers. Throughout the 20th century Cebuano was widely used as a lingua franca in Mindanao and was almost universally known as a second language by those in Mindanao who were not native speakers of Cebuano. At the present time Tagalog is gaining as the lingua franca at the expense of Cebuano, and in Mindanao, as throughout the Cebuano speech area, native speakers of Cebuano are more and more learning Tagalog as a second language. Cebuano is considered a language of the home and social intercourse, and as such enjoys little prestige and is excluded from settings that are considered official or involve people of high rank. For these settings English is used. Further, the educated classes use English as a code together with Cebuano in social settings. Church services that aim at a lower-class audience are in Cebuano, but those aiming at an upper-class congregation are held in English. Books are in English, and English is the official medium of instruction, although for practical reasons teachers make frequent resort to Cebuano at the primary and even secondary levels (the children do not understand English). As an upshot of the emphasis given to English in the educational system and Cebuano's lack of prestige, the elite know the latter but poorly and speak a kind of basic Cebuano mixed with

English, which does not make full use of the rich vocabulary and grammatical apparatus which would allow for eloquence. The best knowledge of Cebuano and most eloquent use is on the part of low-status groups, people with little education and little access to English. Cebuano was widely used in mass media until the middle of the 20th century, but in recent years Tagalog has become more and more widespread. There are still radio programs in Cebuano, and there is one weekly, *Bisaya*, distributed throughout the Cebuano-speaking area, which is aimed at a readership with little education.

Cebuano was first recorded in 1521 in a word list written down by Pigafetta, Magellan's chronicler, when Magellan's expedition made its ill-fated stop in Cebu. Catechisms in Cebuano were composed in the years shortly after the first Spanish colonization in 1564, and the translations made at this time are still in use. The earliest dictionaries and grammatical sketches were composed during the 17th century, although none of these were published until the 18th century. Otherwise no literature antedating the 20th century survives, but the beginning of the 20th century saw a surge of interest in Cebuano and the beginnings of a rich literary production, which gradually diminished from the 1920s and 1930s to the point that now very little is being written. The early dictionaries and catechisms of Cebuano show that the language has changed considerably since the 17th century. Many of the verb forms used in the catechisms and cited in the earliest dictionary are no longer used (although remnants are found in rural dialects) and others are confined to ceremonious or particularly fancy styles, and absent from normal speech. In vocabulary, too, the language has changed considerably. At least one-third of the listings in the major Cebuano dictionary by Fr. Juan Felix de la Encarnación, which dates from the middle of the 17th century, were unknown to more than 100 informants queried during the 1960s and 1970s.

What Cebuano Is Like in Comparison with Tagalog

Cebuano is typologically like the other languages of the Philippines, and most similar to Tagalog (see **Tagalog**). The sound systems of the two languages are similar, but have a very different rhythm, for two reasons. First, Tagalog loses the glottal stop in any position except before pause, whereas Cebuano pronounces the glottal stop with a sharp clear break, giving a staccato effect to the language. Second, Tagalog has short and long vowels, with no limit on the number of long vowels within a word or on

the syllable on which length occurs. Cebuano has few long vowels, and only on the final syllable. The Tagalog and Cebuano consonant inventories are exactly the same. The vowels are different, however. Cebuano has only three vowels, /i/, /a/, and /u/. (Some dialects retain a fourth central vowel, schwa, inherited from Proto-Austronesian, but this has merged with /u/ in the Cebuano of Cebu City.) The vowels /a/ and /u/ may occur lengthened in the final syllable. Stress is contrastive and occurs on the final or the penult. There can be no more than one long vowel in a word.

The Cebuano verb system is similar to Tagalog's but not commensurate with it: the Cebuano verb expresses tense (action started or not), and also has special tenseless forms which are used when the verb is preceded by an adverb or phrase which expresses tense. These three verb forms are durative or non-durative, as exemplified below:

- (1) Action started, punctual vs. action started, durative:

| | | | |
|--------------------------------------|------|-----|------|
| misul?ub | siya | ug | pula |
| put-on | she | OBJ | red |
| 'she put something red on' | | | |
| nagsul?ub | siya | ug | pula |
| is-wearing | she | OBJ | red |
| 'she is (was) wearing something red' | | | |

- (2) Action not started, punctual vs. durative:

| | | | |
|-------------------------------------|------|-----|------|
| musulub | siya | ug | pula |
| put-on | she | OBJ | red |
| 'she will put on something red' | | | |
| magsul?ub | siya | ug | pula |
| is-wearing | she | OBJ | red |
| 'she will be wearing something red' | | | |

- (3) Tenseless verb, durative vs. punctual:

| | | | | |
|-----------------------------------|------|------------|-----|------|
| wa? | siya | musul?ub | ug | pula |
| not | she | put-on | OBJ | red |
| 'she didn't put something red on' | | | | |
| wa? | siya | magsul?ub | ug | pula |
| not | she | is-wearing | OBJ | red |
| 'she wasn't wearing red' | | | | |

A system of affixes which show prepositionlike relationships, analogous to that shown by the Tagalog verb, cuts across this tense-aspect system of Cebuano: the Cebuano verbs contain morphemes which express the relation between the verb and a word it refers to. The verb may refer to the agent (active voice), the patient of the action (direct passive), the thing moved or said (conveyance passive), the instrument of the action, the place of the action, the beneficiary of the action, or (peculiarly for Cebuano) time of the action:

- (4) (Active)

| | | | |
|--|--------|-----|---------|
| Mipalit | siya | ug | ságing |
| bought | he/she | OBJ | bananas |
| 'he bought some bananas [that's what he did]' | | | |

- (5) (Patient)

| | | | |
|--|--------|-----|---------|
| Gipalit | níya | ang | ságing |
| bought-it | by-him | the | bananas |
| 'he bought the bananas [that's what happened to the bananas]' | | | |

- (6) (Place)

| | | | |
|--|-----|---------|------------------|
| bálik | ta | sa | gipalitan |
| let's-go-back | we | to | was-bought-at |
| nímu | ug | ságing | |
| by-you | OBJ | bananas | |
| 'let's go back to the place you bought some bananas' | | | |

- (7) (Instrument)

| | | | |
|---|--------|--------------|---------|
| Ma?u | na | y | |
| is-the-one | that | the-one-that | |
| ipalit | nímu | ug | ságing |
| will-buy-with-it | by-you | OBJ | bananas |
| 'that is the thing [money] you will use to buy bananas with ' | | | |

- (8) (Beneficiary)

| | | | | |
|----------------------------------|--------|-----------------|--------|------|
| Putling | Mariya | ig?ampu? | mu | kami |
| Virgin | Mary | pray-for | by-you | us |
| 'Virgin Mary pray for us' | | | | |

These verbal inflections are added to roots. In addition, new stems can be formed by adding one or more derivational affixes that have meanings similar to those found in Tagalog (see **Tagalog**).

Cebuano has a complex system of deictics and demonstrative pronouns that is a good deal more complex than that of Tagalog. The deictics in Cebuano distinguish tense when initial in the clause: e.g., *dinhi* 'was here', *níla* 'is here', *anhi* 'will be here.' They distinguish for four distances, *díla* 'is here near me (but not near you)', *níla* 'is here (near you and me)', *nála* 'is there (near you but not near me)', *tíla* 'is there (far from both of us)'. When final in the clause the deictics distinguish motion from nonmotion: *didtu* 'there (far away)', *ngadtu* 'going there (far away)'. The interrogatives forms for 'when' and 'where' also distinguish tense.

The changes that Cebuano has undergone since the earliest attestations amount to the loss of distinctions. This can be accounted for partly by the fact that Cebuano has been brought to new areas and spread to populations formerly speaking other languages and also by the fact that there has never been a prescriptive tradition which derogates deviant forms. The four-vowel system, which Cebuano inherited from the protolanguage, has been reduced to

three, except in the case of rural dialects. Further, the category durative vs. punctual, which characterizes the verbal system, has in historical times been lost in the passive verbs except in ceremonial styles. Many of the derivational affixes forming verb stems that were productive in pre-19th-century attestations of the language are now confined to petrified forms. In the past two generations Tagalog has influenced an important component of the verbal system, namely, the loss of the tenseless forms, although in rural speech this part of the system is still intact. Further, the system of deictics has been simplified in speakers influenced by Tagalog: namely, tense has been lost, the four-way distance distinction has been reduced to two – i.e., ‘here’ vs. ‘there,’ and the distinction between deictics expressing motion and those which do not has been lost. These changes are most strongly observed in areas which or among groups who have contact with Tagalog speech, and from this population these simplifications spread elsewhere in the Cebuano speech community.

Cebuano morphology differs in type of Tagalog in two ways: first, affixational patterns are regular and predictable in Tagalog but in Cebuano they are not: whereas in Tagalog the paradigms are normally filled out for all roots with a given meaning

type, in Cebuano many affixes are capriciously distributed, quite irrespective of the semantic qualities of the root. Second, there are numerous variations in affixation and some of the interrogatives, distributed by areas and individual speakers. Tagalog has much less variation.

Bibliography

- Cabonce R (1983). *An English-Cebuano Visayan dictionary*. Manila: National Book Store.
- Encarnación Fr J F de la (1885). *Diccionario Bisaya-Español* (3rd edn.). Manila.
- Mojares R B (1977). *Bibliography of Cebuano linguistics*. Cebu City: University of San Carlos.
- Wolff J U (1961). *Cebuano texts with glossary and grammar*. Mimeographed. Cebu City.
- Wolff J U (1966–7). *Beginning Cebuano* (2 vols). New Haven: Yale University Press.
- Wolff J U (1972). *A Cebuano Visayan dictionary*. Ithaca, NY: Southeast Asia Program, Cornell University/Manila: Linguistic Society of the Philippines 72–81.
- Wolff J U (1973). ‘The character of borrowings from Spanish and English in the languages of the Philippines.’ *Journal of Philippine Linguistics* 4(1).
- Zorc D (1977). *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.

Celtic

C Ó Baoill, University of Aberdeen, Aberdeen, UK

© 2006 Elsevier Ltd. All rights reserved.

The Celts get their name from *Keltoi*, a name of unknown origin applied by the Greeks from around 500 B.C. to a widespread people who lived mainly to the north and west of them. They have long been identified with the archaeological cultures known as Hallstatt and La Tène, named from type-sites in central Europe and dating from the period following 600 B.C., but linking a language to an archaeological culture can be unreliable, and this link and others concerned with the Celts have been queried, notably in James (1999).

The languages understood to belong to these people are of the Indo-European family, the most westerly branch of it, and one important feature thought to mark Celtic out from the rest is the loss (or reduction in some contexts) of the letter *p*. For example, the Indo-European word for a ‘father,’

which began with *p*- (whence, e.g., Greek and Latin *pater*), gives modern Gaelic (Gaelic, Irish) *athair*. This development predates all the evidence we have for the languages. Another early development was the change in some branches of Celtic, whereby the Indo-European /k^u/ (or ‘Q’) became /p/, whence the well-known division between P-Celtic and Q-Celtic languages. In the later (insular Q-Celtic) languages this *q* has developed to a /k/ sound, written *c*, and so we get oppositions like Gaelic *cenn* and Welsh *pen*, ‘head’ (from an original stem **qen*-).

The languages may be classified as Continental Celtic and Insular Celtic, the former group dating from the earliest period of Celtic history up till about 500 A.D., by which time all the continental languages had probably disappeared. Three main continental languages are identifiable, Gaulish, Lepontic, and Celtiberian, and we know all three principally from inscriptions (on stones or on coins), names (place-names and personal names) and quotations on record in other languages. Verbs, and therefore sentences, are extremely rare, so that our

knowledge of all three languages really is minimal. Gaulish and Lepontic are P-Celtic languages, the former belonging to the general area of Gaul (France, but including also parts of Switzerland, Belgium, and Italy) and the latter to parts of the southern Alps. Celtiberian is the name favored, over the alternative Hispano-Celtic, by de Hoz (1988) for the Q-Celtic language, which has, since the mid-20th century, come to be reasonably well attested by inscriptions in north central Spain; a relevant opposition here is between the form used for 'and' (Latin *-que*), appearing as *pe* in Lepontic and as *cue* in Celtiberian.

Archaeology indicates movement of features of the Hallstatt and La Tène cultures from the continent to Britain and Ireland from about 500 B.C., and it is assumed that Celtic languages came with them. Jackson (1953: 4) used the term Gallo-Brittonic to cover both Gaulish and the first P-Celtic languages in Britain. A Q-Celtic language appeared in Ireland, but there is much disagreement as to when, whence, and by what route. There is also much discussion of criteria for assessing relationships between the Celtic languages in this early period, and opinions change frequently (see Evans, 1995); evidence for dating expansion and change in the languages is inevitably scarce.

The Insular Celtic languages are divided into Brythonic and Goidelic groups, the former denoting the descendants of the P-Celtic, which reached Britain from the continent, namely Welsh, Cornish, Breton, Pictish, and Cumbric. Cumbric (or Cumbrian) is used to denote the early language(s) of what are now the northern part of England and the southern part of Scotland, but little is really known about the language(s) apart from what can be gathered from names (see Price 1984: 146–154). The surviving languages in the Brythonic group are Welsh and Breton, Cornish having gone out of general use in the 18th century, though it is still in use among enthusiasts. Sims-Williams (1990: 260; see also Russell, 1995: 132–134) argued that the main linguistic developments from (the theoretical) Brittonic, leading toward the modern insular languages, were in place by 500 A.D., and divergences between Cornish and Breton followed shortly afterward.

Goidelic is the term used by linguists for the Q-Celtic language that appeared in Ireland before the 1st century B.C. and for its descendants. The theory has long been that the original Goidelic language in Ireland spread to western Britain when the power of the Romans waned around 400 A.D., and that Scottish Gaelic (Gaelic, Scots) and Manx eventually developed there. But while the simple theory of a major Irish migration bringing Gaelic to Scotland is widely accepted, even in Scotland, Ewan Campbell

has recently shown (Campbell, 2001) that archaeology provides no evidence in support of any such invasion.

The earliest written form of the Gaelic language is that found in Ogam, the alphabet used for inscriptions on stone, dating from about the 4th century till the 7th (McManus, 1991 is a detailed study). Thereafter the language, as attested in the literature, is divided into Old (till 900 A.D.), Middle (900–1200), Early Modern (till c. 1650), and Modern periods. The distinctive Scottish and Manx forms only become clearly visible in the Early Modern period. The linguistic theory in Jackson (1951: 78–93) envisaged a historical period, c. 1000–1300 A.D., during which Irish (as Western Gaelic) became clearly distinct from Eastern Gaelic (Scottish Gaelic and Manx), but this has come under attack by those (such as Ó Buachalla, 2002) who see the significant historical division within Goidelic as a north/south one, with Scotland, Man, and Ulster in opposition to the rest of Ireland on many points.

On similar grounds, the three Gaelic languages may be seen rather as what Hockett (1958: 323–325) called an L-complex, a single linguistic continuum within which national and even geographical boundaries are ignored by dialectal isoglosses. This suggestion (cf. Ó Buachalla, 1977: 95–96) is supported (a) by the fact that all three 'languages' identify themselves by variants of the same name, *Gaeilge*, *Gàidhlig*, *Gaelck*, and others, whence the English term Gaelic; and (b) by the strong evidence that, while Gaelic survived (until the early 20th century) in the interface area between north-eastern Ireland and the southern Highlands, speakers on both sides of the North Channel were able to converse with little difficulty.

Bibliography

- Campbell E (2001). 'Were the Scots Irish?' *Antiquity* 75, 285–292.
- de Hoz J (1988). 'Hispano-Celtic and Celtiberian.' In MacIannan G (ed.) *Proceedings of the First North American Congress of Celtic Studies*. Ottawa: University of Ottawa. 191–207.
- Evans D E (1995). 'The early Celts: the evidence of language.' In Green M J (ed.) *The Celtic world*. London: Routledge. 8–20.
- Hockett C F (1958). *A course in modern linguistics*. New York: Macmillan.
- Jackson K (1951). '“Common Gaelic”: the evolution of the Goidelic languages.' In *Proceedings of the British Academy XXXVII*, 71–97.
- Jackson K (1953). *Language and history in early Britain*. Cambridge: Cambridge University Press.

- James S (1999). *The Atlantic Celts: ancient people or modern invention?* London: British Museum Press.
- Maier B (2003). *The Celts: a history from earliest times to the present*. Edinburgh: Edinburgh University Press.
- McManus D (1991). *A guide to Ogam*. Maynooth: An Sagart.
- Ó Buachalla B (1977). 'Ní and cha in Ulster Irish' *Ériu* 28, 92–141.
- Ó Buachalla B (2002). '“Common Gaelic” revisited.' In Ó Baoill C & McGuire N R (eds.) *Ramnsachadh na Gàidhlig 2000*. Obar Dheathain: An Clò Gaidhealach. 1–12.
- Price G (1984). *The languages of Britain*. London: E. Arnold.
- Russell P (1995). *An introduction to the Celtic languages*. London: Longman.
- Sims-Williams P (1990). 'Dating the transition to Neo-Brittonic: phonology and history, 400–600.' In Bammesberger A & Wollmann A (eds.) *Britain 400–600: language and history*. Heidelberg: C. Winter. 217–261.

Central Siberian Yupik as a Polysynthetic Language

W J de Reuse, University of North Texas, Denton, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

An Overview of the Central Siberian Yupik Word

Central Siberian Yupik (CSY) is a representative language of the Yupik branch of the Eskimo-Aleut family. It is spoken by over 1000 people on St. Lawrence Island, Alaska and Chukotka, Russian Far East (de Reuse, 1994; Nagai, 2004). Like all Eskimo languages, CSY is, from a typological point of view, extreme because of its high level of polysynthesis, and the fact that it is almost exclusively suffixing (Woodbury, 2002: 98). There is no compounding, and CSY has only one prefix, occurring as a lexicalized element on demonstratives. The structure of the Eskimo noun or verb word can be schematized as follows:

- (1) Base + postbasesⁿ + ending + enclitic^m

The base is the lexical core of the word; it can be followed by a number *n* of postbases. The value of *n* is between 0 and a theoretically infinite number, but *n* > 6 is quite rare. Postbases are traditionally considered derivational suffixes and combine with the base to form a new base. The obligatory ending is inflectional, marking case, number, and possession for nouns, marking mood, person, and number of subject for intransitive verbs, and marking mood, person and number of subject and person and number of object for transitive verbs. Although there are about 1200 inflectional endings for ordinary verbs (Woodbury, 2002: 81), it is not the richness of inflection that characterizes CSY as a polysynthetic language, since its inflection is not very different from that found in Latin or Ancient Greek. Enclitics, of which there are 12, can follow the ending. They are

syntactic particles that form a phonological word with the immediately preceding word. The value of *m* is between 0 and 4. Example (2) is an analysis of a CSY word that illustrates the structure in Schematic (1) (abbreviations: V, verb; PST, past tense; FRUSTR, frustrative ('but . . . , in vain'); INFER, inferential evidential (often translatable as 'it turns out'); INDIC, indicative; 3s.3s, third-person subject acting on third-person object):

- (2) neghyaghtughyugumayaghpetaaallu
 negh- -yaghtugh- -yug- -uma-
 eat go.to.V want.to.V PST
 -yagh- -pete- -aa =llu
 FRUSTR INFER INDIC.3s.3s also
 'Also, it turns out she/he wanted to go eat it, but . . . '

In Example (2), only the base *negh-* and the inflectional *-aa*, are obligatory. Any or all of the other suffixes, which are postbases, can be left out. The element = *llu* is an enclitic.

Polysynthesis Illustrated by CSY Postbases

Since the postbases account for the polysynthesis of CSY, we will focus on their characteristics. A first characteristic is the full productivity of most (but not all) postbases. The five postbases of Example (2) are fully productive. So, picking between one and five postbases from the five in Example (2), it is possible to generate 30 different words. For semantic reasons, it happens to be the case that the order of elements has to be *-yaghtugh-yug-uma-yagh-pete-*. There are no clear morphological position classes to be set up in CSY. A second characteristic of some CSY postbases is recursion, as illustrated by Example (3):

- (3) iitghesqesaghiisqaa
 itegh- -sqe- -yaghtugh-
 come.in ask.to.V go.to.V

-sqe- -aa
ask.to.v INDIC.3s.3s
'He_i asked him_j to go ask him_k to come in'.

The postbase *sqe-* 'ask to.v' is used recursively. A third characteristic of some CSY postbases is that they can display variable order with respect to each other without resulting differences in meaning. This is illustrated with Examples (4) and (5):

- (4) aananiitkaa
aane- -nanigh- -utke-
go.out cease.to.v v.on.account.of
-aa
INDIC.3s.3s
'He ceased going out on account of it'.
- (5) aanutkenanighaa
aane- -utke- -nanigh-
go.out v.on.account.of cease.to.v
-aa
INDIC.3s.3s
'He ceased going out on account of it'.

Even though generally in CSY the rightmost postbase has scope over what is on the left, that principle does not seem to be working in Examples (4) and (5). These two sentences mean exactly the same thing and were uttered within three lines of each other in a story (de Reuse, 1994: 93). A fourth characteristic of postbases is that they can interact with the syntax, and attach to elements functioning as independent syntactic atoms. This is illustrated in Example (6) (abbreviations: ABS, absolutive; 2s.s, second-person singular possessor, singular possessum; INTRANS, intransitive; PARTL, participial mood (often nominalizing in Eskimo); ABL, ablative; N, noun; 3s, third-person singular subject):

- (6) Atan aangelghiimeng qikmilguuq.
ata- -n aange- -lghii-
father ABS.2s.s be.big INTRANS.PARTL
-meng qikmigh- -lgu- -uq
ABL.S dog have.N INDIC.3s
'Your father has a big dog'.

As Sadock (1980, 1991) demonstrated on the basis of parallel structures in Greenlandic Eskimo, the noun-incorporating postbase *-lgu* 'have.N' acts like a morphologically intransitive verb, and like other intransitive verbs, it can occur with a direct object in an oblique case (here the ABL). Since postbases cannot attach to inflected words, the ABL case marking cannot occur on *qikmigh-* 'dog,' but it does show up in the stranded modifier *aangelghiimeng* 'big.' This is expected, since CSY modifiers agree in case with their heads. At the syntactic level then, *aangelghiimeng qikmigh-* 'big dog' forms a phrasal constituent to which the *-lgu-* is attached.

A fifth characteristic of postbases is that they not only derive verbs from verbs (as in Examples (2)–(5)), or nouns from nouns (shown in Example (7)), but also verbs from nouns, as in Example (6), and nouns from verbs, as in Example (7). This is, of course, expected behavior for derivational morphology. Example (7) contains the verb *yughagh-* 'to pray', changing to a noun *yughaghvig-* 'church', changing to another noun *yughaghvigllag-* 'big church', and changing back to a verb *yughaghvigllange-* 'to acquire a big church' (abbreviation: 3P, third-person plural subject).

- (7) yughaghvigllanguyutut
yughagh- -vig- -ghllag- -nge-
pray place.to.v big.N acquire.N
-yug- -tut
want. to.v INDIC.3P
'They want to acquire a big church.'

As noted earlier, not all postbases are productive. The postbase *-vig-* 'place to.v' is an example of a nonproductive postbase, since it lexicalized with 'pray' to mean 'church,' and not the completely predictable 'place to pray,' i.e., any place to pray. The postbases that follow *-vig-* are completely productive. There are over 400 productive postbases in CSY, and several hundred nonproductive ones.

Productive Postbases: Neither Derivation nor Inflection?

The survey of the characteristics of productive postbases just provided casts some doubt on their status as elements of derivational morphology. Certainly, the nonproductive postbases behave like elements of derivational morphology. Regarding productive postbases, consider Table 1, a chart of criteria distinguishing inflection, (nonproductive) derivation, productive postbases, and syntax. The productive postbases, even though bound, have six features in common with syntax; they also have one (feature [6]) in common with derivation, and two (features [1] and [5]) in common with inflection. In the following explanations, the term 'elements' will be used instead of 'productive postbase' or 'words,' in order to have a term covering both morphology and syntax. The criteria of the six features are intended to show that elements such as productive postbases are syntax-like. Presumably the criteria in Table 1 are not independent of each other, but it is not yet clear which has to be derived from which.

Productivity (feature [1]) means that there are no idiosyncratic restrictions on the use of the element. Thus, its presence is conditioned by semantic plausibility only, and not by selectional restrictions.

Table 1 Criteria of inflection, derivation, productive postbases, and syntax

| Feature | Inflection | (Nonproductive) derivation | Productive postbases | Syntax |
|--|------------|-------------------------------|-------------------------|--------|
| [1] Productive? | Yes | No | Yes | Yes |
| [2] Recursion possible? | No | No | Yes | Yes |
| [3] Necessarily concatenative? | No | No | Yes | Yes |
| [4] Variable order of elements possible in some instances? | No | No | Yes | Yes |
| [5] Interaction with syntax possible? | Yes | No | Yes | Yes |
| [6] Lexical category changing possible? | No | Yes | Yes | Yes |

Certainly in CSY, and for many polysynthetic languages, the elements are so numerous that it is very unlikely that native speakers would have the ability to memorize the existing sequences and store them in the lexicon (Fortescue, 1980; de Reuse, 1994). Inflection, of course, is also completely productive, but only within a paradigm. The claim is that derivational morphology is never fully productive. Since some of what is traditionally called ‘derivational morphology’ is productive, we are, in effect, changing the definition of derivational morphology, so that fully productive elements of derivational morphology are no longer part of it.

Recursion (feature [2]) means that the same element can potentially occur more than once within the same word (which is the case with productive postbases), or within the same sentence (which is the case in syntax), its presence again conditioned by semantic plausibility.

Concatenative (feature [3]) means that the elements are going to be in some linear order. Neither nonconcatenative morphology, such as suppletion, nor Semitic style morpheme internal change is expected to exist instead of postbases. Similarly, nonconcatenative syntax does not exist.

Variable order (feature [4]) means that, in some cases, the order of elements can be free. Just as in free word order in syntax, some productive postbases can be freely ordered, most likely constrained by pragmatic factors only. This is impossible in derivation.

Interaction with syntax (feature [5]) has to do with relationships between the productive postbases and elements of syntax. As is well known (Anderson, 1982), inflection interacts with syntax, as in agreement or case marking. Derivation does not interact with syntax, but productive postbases do interact with syntax. And obviously, syntax interacts with itself.

Lexical category changing (feature [6]) means that the element can change the lexical category in the morphology. Derivational morphology can do this, but inflectional morphology does not. Here, postbases behave like derivational morphology. In a

parallel fashion, in the syntax, the addition of an element can change the phrasal category. For example, *very good* is an adjective phrase, but *very good quality* is a noun phrase.

These characteristics of Eskimo productive postbases lead us to suggest the existence of a branch of morphology, which is neither inflection, nor derivation, that we will call ‘productive noninflectional concatenation,’ or PNC (PNC was called ‘internal syntax’ in de Reuse (1992)). The term ‘concatenation,’ rather than ‘affixation,’ is used to highlight the fact that PNC can be affixal (as in Eskimo) or compounding. It is proposed that the existence of large amounts of PNC elements is a valid way of characterizing polysynthetic languages.

Consequences for a Productive Noninflectional Concatenation View of Polysynthesis for Morphological Theory

The proposal that polysynthesis can be characterized in terms of PNC has consequences for morphological theory. If it is assumed, for example, that productivity is definitional of PNC, it is necessary to account for productive affixation in nonpolysynthetic languages. Indeed, some of the affixes traditionally called derivational in Indo-European languages are completely productive, and among these productive ones, some are recursive as well. Examples of productive and recursive prefixes in English are *anti-*, as in *antiabortion*, *antiantiabortion*, etc., or, more marginally, *re-*, as in *rewrite*, *rerewrite*, etc. The diminutive suffix of Dutch, *-je*, is completely productive. The diminutive of Dutch contrasts starkly with the diminutive suffixes of French (*-et*, *-ette*), and the diminutive suffixes of English (*-ette*, *-let*, *-kin*, *-ling*), which are unproductive. As a result, *anti-*, *re-*, and the Dutch diminutive must be considered to be PNC elements, rather than derivational ones. The difference with polysynthetic languages is a quantitative one. European languages have just a few elements of PNC. Mildly polysynthetic languages (such as found in the Arawakan and Siouan families) have more than a dozen of such elements, solidly polysynthetic languages (such as found in the

Caddoan and Wakashan families) have over 100 of such elements, and extreme polysynthetic languages (i.e., the Eskimo branch of Eskimo-Aleut) have several hundreds of such elements.

Within polysynthetic languages, it will also be necessary to distinguish between their nonproductive morphology (derivation or compounding) and PNC. According to Mithun and Gorbett's research (1999) on noun incorporation in Iroquoian, speakers can often tell which combinations are being used and which ones are not being used. If that is so, some of the noun-incorporating morphology of Iroquoian is not productive, and should not count for considering the language polysynthetic. Similarly, a distinction must be made, in Eskimo, between nonproductive postbases, such as *-vig-* 'place to.v,' as in Example (7), which do not count for considering the language polysynthetic, and the elements of PNC, i.e., the productive postbases, for which the question of which combinations are used or not used cannot be reasonably answered.

Bibliography

Anderson S R (1982). 'Where's morphology?' *Linguistic Inquiry* 13, 571–612.

de Reuse W J (1992). 'The role of internal syntax in the historical morphology of Eskimo.' In Aronoff M (ed.) *Morphology now*. Albany: State University of New York Press. 163–178.

de Reuse W J (1994). *Siberian Yupik Eskimo. The language and its contacts with Chukchi*. Salt Lake City: University of Utah Press.

Fortescue M (1980). 'Affix-ordering in West Greenlandic derivational processes.' *International Journal of American Linguistics* 46, 259–278.

Mithun M & Gorbett G (1999). 'The effect of noun incorporation on argument structure.' In Mereu L (ed.) *Boundaries of morphology and syntax*. Amsterdam: John Benjamins. 49–71.

Nagai K (2004). A morphological study of St. Lawrence Island Yupik: three topics on referentiality. Ph.D. diss. (linguistics), Kyoto University.

Sadock J M (1980). 'Noun incorporation in Greenlandic: a case of syntactic word-formation.' *Language* 57, 300–319.

Sadock J M (1991). *Autolexical syntax. A theory of parallel grammatical representations*. Chicago & London: The University of Chicago Press.

Woodbury A C (2002). 'The word in Cup'ik.' In Dixon R M W & Aikhenvald A Y (eds.) *Word. A cross-linguistic typology*. Cambridge: Cambridge University Press. 79–99.

Central Solomon Languages

A Terrill, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

There are four or possibly five Papuan languages in the central Solomon Islands: Bilua, spoken on the island of Vella Lavella; Touo (known more commonly in the literature as Baniata, after one of the villages where it is spoken), spoken on Rendova Island; Lavukaleve, spoken in the Russell Islands; Savosavo, spoken on Savo Island; and possibly Kazukuru, an extinct and barely documented language of New Georgia.

Relationships Among the Languages

By the time of Ray (1926, 1928), there was already an established list of non-Austronesian languages of the Solomon Islands, consisting of Bilua, Baniata (here referred to as Touo), Savo, and Laumbe (now called Lavukaleve). Waterhouse and Ray (1931) later discovered Kazukuru, a language of New Georgia,

identifying it as unlike both the Melanesian (i.e., Austronesian) and Papuan languages of the Solomon Islands. Much later, Lanyon-Orgill (1953) claimed Kazukuru and two further varieties, Guliguli and Dororo, to be Papuan languages; however, the data are so scant as to make classification uncertain.

Greenberg (1971) was the first to make an explicit claim for the genetic unity of these languages, as part of his Indo-Pacific family. This claim was shortly followed by Wurm's (1972, 1975, 1982) proposal of an East Papuan phylum, linking all the Papuan languages of the islands off the coast of New Guinea into one genetic grouping. Both claims have been firmly rejected by specialists in the region, and recent views have been much more cautious: Ross (2001) suggested, on the basis of similarities in pronouns, that Bilua, Touo (Baniata), Savosavo, and Lavukaleve formed a family, unrelated to other island and mainland Papuan languages. Terrill (2002) found limited evidence of similarities in gender morphology among these languages. In lexical comparisons using an extended Swadesh list of roughly 333 items (with obvious Austronesian loans removed), Bilua, Lavukaleve,

Touo, and Savosavo share only 3–5% resemblant forms (i.e., within the realm of chance). In short, at this stage of knowledge, a genetic relationship among any or all of these languages still remains to be proven.

Typological Characteristics

A typological overview of these and other Papuan languages of island Melanesia provided by Dunn *et al.* (2002) showed that, but for a few striking exceptions, the only grammatical features shared by the central Solomon Islands Papuan languages are also held in common with surrounding Oceanic Austronesian languages. These common features include an inclusive/exclusive distinction in pronouns, dual number (actually, there are four number categories in Touo), reduplication for various purposes, nominative/accusative alignment (although Lavukaleve has ergative/absolutive alignment in certain types of subordinate clauses), and serial verb constructions (absent in Bilua).

The two most notable departures from Oceanic grammatical patterns are SOV constituent order in three of the languages (Bilua has SVO with some variation) and the presence of gender; there are three genders in Lavukaleve, four in Touo, and two in Bilua and Savosavo. Gender in Bilua is contextually determined: the masculine–feminine distinction applies only to human nouns, but for inanimate nouns there is a distinction, marked by the same morphology as marks gender in human nouns, between ‘singulative’ (=masculine) and ‘unspecified number’ (=feminine) (Obata, 2003). Savosavo has two genders, masculine and feminine, and it is not clear whether they are contextually determined as in Bilua or permanently assigned as in Touo and Lavukaleve (Todd, 1975).

Touo has some very unusual features for the region, including a phonological distinction between breathy/creaky vs. modal vowels, as well as six vowel positions instead of the usual five for the region. Touo sources include Todd (1975), Frahm (1999), and Terrill and Dunn (2003). Lavukaleve too has many unusual features, including focus markers that show agreement in person, gender, and number of the head of the constituent on which they mark focus; and a very complex participant marking system depending on factors to do with predicate type and clause type (Terrill, 2003).

Bibliography

- Dunn M, Reesink G & Terrill A (2002). ‘The East Papuan languages: a preliminary typological appraisal.’ *Oceanic Linguistics* 41, 28–62.
- Frahm R M (1999). Baniata serial verb constructions, M.A. thesis, University of Auckland.
- Greenberg J H (1971). ‘The Indo-Pacific hypothesis.’ In Sebeok T A (ed.) *Current trends in linguistics, vol. 8: Linguistics in Oceania*. The Hague: Mouton and Co. 807–871.
- Lanyon-Orgill P A (1953). ‘The Papuan languages of the New Georgian Archipelago, Solomon Islands.’ *Journal of Austronesian Studies* 1, 122–138.
- Obata K (2003). *A grammar of Bilua: a Papuan language of the Solomon Islands*. Canberra: Pacific Linguistics 540.
- Ray S H (1926). *A comparative study of the Melanesian Island languages*. London: Cambridge University Press.
- Ray S H (1928). ‘The non-Melanesian languages of the Solomon Islands.’ In Koppers W (ed.) *Festschrift publication d’hommage offerte au P. W. Schmidt*. Vienna: Mechitharisten-Congregations-Buchdruckerei. 123–126.
- Ross M (2001). ‘Is there an East Papuan phylum? Evidence from pronouns.’ In Pawley A, Ross M & Tryon D (eds.) *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*. Canberra: Pacific Linguistics. 301–321.
- Terrill A (2002). ‘Systems of nominal classification in East Papuan languages.’ *Oceanic Linguistics* 41, 63–88.
- Terrill A (2003). *A grammar of Lavukaleve*. Berlin: Mouton de Gruyter.
- Terrill A & Dunn M (2003). ‘Orthographic design in the Solomon Islands: the social, historical, and linguistic situation of Touo (Baniata).’ *Written Languages and Literacy* 6, 177–192.
- Todd E (1975). ‘The Solomon Language family.’ In Wurm S A (ed.) *Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics C-38. 805–846.
- Waterhouse W H L & Ray S H (1931). ‘The Kazukuru language of New Georgia.’ *Man* xxxi, 123–126.
- Wurm S A (1972). ‘The classification of Papuan languages and its problems.’ *Linguistic Communications* 6, 118–178.
- Wurm S A (1975). ‘The East Papuan phylum in general.’ In Wurm S A (ed.) *Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics C-38. 783–804.
- Wurm S A (1982). *Papuan languages of Oceania*. Tübingen: Gunter Narr Verlag.

Chadic Languages

P J Jaggar, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Chadic language family comprises an estimated 140 to 150 languages spoken in areas to the west, south, and east of Lake Chad (west Africa). The best-known and most widespread Chadic language is Hausa, with upwards of 30 million first-language speakers, more than any other language in Africa south of the Sahara. The remaining languages, some of which are rapidly dying out (often due to pressure from Hausa), probably number little more than several million speakers in total, varying in size from fewer than half a million to just a handful of speakers, and new languages continue to be reported. Written descriptions of varying length and quality are available for only about one-third of the total, although for some – e.g., Bidiya (Bidiyo), Guruntum, Kanakuru (Dera), Kera, Kwami, Lamang, Margi (Marghi Central), Miya, and Mupun – good descriptive grammars have been produced, and several dictionaries have appeared, e.g., Dangaléat, Lamé, Ngizim, and Tangale. Hausa has four recent comprehensive reference grammars, in addition to two high-quality dictionaries, making it the best-documented language in sub-Saharan Africa.

Chadic is a constituent of the Afroasiatic phylum, which also includes Semitic (e.g., Amharic, Arabic, [Standard] Hebrew), Cushitic (e.g., Oromo, Somali), Omotic (e.g., Dime, Wolaytta), Berber (e.g., Tamahaq and Tamajeq [Tamajeq, Tayart] [spoken by the Tuareg], Tamazight [Central Atlas], and (extinct) Ancient Egyptian/Coptic. The phylogenetic membership of Chadic within Afroasiatic was first proposed almost 150 years ago, but did not receive wide acceptance until Greenberg's (1963) major (re)classification of African languages. The standard internal classification divides Chadic languages into three major branches: West (e.g., Hausa, Bole, Angas, Ron, Bade), Central = Biu-Mandara (e.g., Tera, Mandara, Bachama-Bata [Bacama], Kotoko [Afade]), and East (e.g., Somrai, Kera, Dangaléat), in addition to an isolated Masa cluster (with subbranches and smaller groupings).

Phonology

Laryngealized implosive stops, e.g., /b dʒ/, and ejective stops, e.g., /p' t'/, are widespread throughout Chadic, together with prenasalized obstruents, e.g., /mb nd/. A characteristic pattern, therefore, is for a

language to present a four-way phonation contrast, e.g., coronal /t d d' nd/ and/or labial /p b b' mb/. The voiceless and voiced lateral fricatives /ɬ ɮ/ are also commonplace, in addition to palatal and velar (including labialized velar) consonants.

Vowel systems generally vary from two (monophthongal) vowels, high /ə/ (with various phonetic values) and low /a/, as in Bachama-Bata and Mandara, to seven vowels, e.g., [Dangaléat] /i e ε a ɔ o u/, with /i (e) a ə (o) u/ a common inventory, and the diphthongs /ai/ and /au/ are attested. Tangale has a nine-vowel ATR pattern. Contrastive vowel length, especially in medial position, is also widespread throughout the family.

Chadic languages are tonal, and two level (High/Low) tones, e.g., Hausa, or three (High/Mid/Low), e.g., Angas, are typical. Downstep is also common (e.g., Ga'anda, Miya, Tera). Although tone can be lexically contrastive, its primary function is normally grammatical, e.g., in distinguishing tense/aspect/mood categories. [Transcription: aa = long vowel, a = short; â(a) = L(ow) tone, â(a) = F(alling) tone, H(igh) tone is unmarked.]

Morphology and Syntax

Many Chadic languages have masculine/feminine grammatical gender (an inherited Afroasiatic feature), with no distinction in the plural, and typically distinguish gender in second and third person singular pronouns, e.g., [Miya] *fiy/macə* 'you (MASC/FEM)', *tə/njə* 'he/she'. Some also preserve the characteristic n/t/n (MASC/FEM/PL) marking pattern in grammatical formatives (and the masculine and plural markers often fall together phonologically), cf., [Masa] *vèt-na* 'rabbit', *vèt-ta* 'female rabbit', *vèdai-na* 'rabbits'.

Noun pluralization is complex, and some widespread plural suffixes are reconstructable for Proto-Chadic, e.g., *-Vn, *-aki, *-i, and *-ai. Examples: (-Vn) *kùmən/kùmənən* 'mouse/mice' [Bade], *miyò/mishən* 'co-wife/co-wives' [Kanakuru], (-aki) *goonaa/gòonàkii* 'farm(s)' [Hausa], (-i) *duwimà/dùwimi* 'guinea fowl(s)' [Gera], (-ai) *mùtù/mutai* 'sore(s)' [Dangaléat]. Other plurals entail infixation of internal -a-, e.g., [Ron] *sàkur/sakwàar* 'leg(s)'. Some languages restrict overt plural marking to a narrow range of nouns (typically humans and animals).

Verbs in many Chadic languages have retained the lexically arbitrary Proto-Chadic distinction between final -a and final -ə verbs (where the final schwa vowel is often pronounced as [i], [ə], or [u]), cf., [Tera] *na* 'see' and *dlə* 'get', [Guruntum] *daa*

'sit' and shi 'eat'. Verbal semantics and valency are modified by the addition of one or more derivational extensions (often fused suffixes). These extensions encode such notions as action in the direction of (centripetal) or away from (centrifugal) a deictic center (often the speaker), or action partially or totally completed, e.g., (totality) sà-nyà 'drink up' < sà 'drink' [Margi]. Some extensions also have a syntactic function, denoting, *inter alia*, transitivity or perfectivity, e.g., (transitivity) yàw-tu 'take down' < yàwwu 'go down' [Bole], kàta-naa 'return' (TRANS) < kàtee 'return' (INTRANS) [Ngizim]. Verb stems can be overtly inflected for tense-aspect-mood by segmental and/or tone changes.

Many languages also have so-called 'pluractional' verbs, which express an action repeated many times or affecting a plurality of subjects (if intransitive) or objects (if transitive), and are formed via prefixal reduplication, ablaut or gemination, e.g., [Guruntum] pàni/pàppàni 'take', [Angas] fwín/fwàn 'untie', [Péro] lofò/lofò 'beat'. In some languages, pluractional stems occur with plural subjects of intransitive verbs and plural objects of transitive verbs, producing ergative-type agreement. In a number of languages, intransitive verbs are followed by an 'intransitive copy pronoun', which maps the person, number, and gender of the coreferential subject, e.g., [Kanakuru] nà pòrò-*no* 'I went out' (literally I went out-*I*).

Derivational and inflectional reduplication is widespread throughout the family (often signaling semantic intensification), ranging from (a) copying of a single segment, e.g., [Miya] pluractional verb tlyaadà 'to hoe repeatedly' < tlyadà 'to hoe', [Bidiya] tàttuk 'very large' < tàtuk 'large'; (b) reduplication of a syllable, e.g., [Hausa] prefixal reduplication of the initial CVC syllable of a sensory noun to form an intensive sensory adjective, as in zùzzurfaa 'very deep' (< zur-zurf-aa) < zurfii 'depth' (with gemination/assimilation of the coda /r/); (c) full reduplication (exact copy), e.g., [Guruntum] kìnì-kìnì 'just like this' < kìnì 'like this', [Kwami] kayò-kayò 'a gallop' < kayò 'a ride', [Tangale] sàñ-sàñ 'very bright' < sàñ 'bright', [Margi] pərda-pərda 'sinewy piece of meat' < pərda 'sinew'.

Like many African languages, Chadic languages often have a lexically autonomous class of highly expressive, phonosemantic words known as 'ideophones'. Ideophones usually pattern syntactically with adverbials and often have their own distinct phonological and phonotactic properties. They typically reinforce the manner of an action, event, or state, e.g., [Ngizim] ɓarak 'with a popping sound', [Miya] ɓakù-ɓakù 'hopping along', [Kwami] (adjectival) dükùdù 'small and broad', [Hausa] kwàngaràm

'with a clang', [Margi] dzùl-dzùl 'jumping high in running (animal)', [Bidiya] ɓorok (Mid tones) 'emphasizes quickness'.

Word order is normally S[ubject] V[erb] O[bject], although VSO order is found in a few Central Chadic/Biu-Mandara languages spoken in the Nigeria-Cameroon border area. Pronominal indirect objects (recipients/goals) are typically realized as verb clitics, whereas nominal indirect objects occur as prepositional phrases to the right of the direct object/theme (Chadic languages are prepositional), cf., [Kanakuru] à jò6-rò landài 'he washed the robe for her' (literally he washed-*for her* robe), and à jò6è landài gèn tamno 'he washed the robe for the woman' (literally he washed robe *for woman*). *Wh*-questions, focus, and relativization usually pattern together in terms of their formal morphosyntactic reflexes, with overt movement, often to left periphery, and special (focus) marking on the inflectional element, e.g., [Hausa] yaarònkà mukà ganii 'it's your boy (that) we saw' (literally boy.your 1pl.FOCUS.PERF see). Some languages allow (or require) in situ *wh*- and (pragmatic) focus constituents, e.g., [Duwai] Saaku ɓàné mù? 'what did Saku cook?' (literally Saku cooked *what*).

Negation in Chadic is typically signaled with a single marker in sentence-final position, e.g., [Guruntum] tãa kyur shau dà 'she will not cook the food' (literally she will cook food NEG), [Kera] wə gùsnà hàrga bà 'he didn't buy her a goat' (literally he bought her goat NEG), sometimes reinforced by an additional pre-verbal negative marker. Comparatives are normally ditransitive constructions with the lexical verb 'exceed, surpass, be more than', i.e., exceed object X in relation to manner Y.

In noun phrase syntax, the normative order for constituents is head-initial, i.e., head noun followed by definite determiners, possessives, numerals, relative clauses, etc. The linear order in genitive constructions is possessee X (+ 'of' linker) + possessor Y, e.g., [Margi] tagu gə Haman 'Haman's horse' (literally horse of Haman). Many Chadic languages also make an overt distinction between alienable and inalienable possession whereby inalienable possession is expressed by direct juxtaposition (i.e., with no overt linker), cf. (inalienable) mōnda Miyim 'Miyim's wife' (literally wife Miyim), and (alienable) gam ma tamnoi 'the woman's ram' (literally ram of woman) [Kanakuru]. Reflexive pronouns and reciprocals (phrasal anaphors) are typically formed with the body-part nouns 'head' and 'body' respectively, e.g., [Kwami] kuu-nì 'himself' (literally head-his), [Miya] tuwatùw-àamà 'each other (we)' (literally body-our).

Bibliography

- Alio K (1986). *Essai de description de la langue bidiya du Guéra (Tchad). Phonologie – Grammaire*. Marburger Studien zur Afrika- und Asienkunde, Serie A, Afrika, 45. Berlin: Dietrich Reimer.
- Ebert K H (1979). *Sprache und Tradition der Kera (Tschad), Teil 3: Grammatik*. Marburger Studien zur Afrika- und Asienkunde, Serie A, Afrika, 15. Berlin: Dietrich Reimer.
- Fédry J (1971). *Dictionnaire dangaléat (Tchad)*. Paris: Afrique et Langage.
- Frajzyngier Z (1993). *A grammar of Mupun*. Sprache und Oralität in Afrika, 14. Berlin: Dietrich Reimer.
- Frajzyngier Z (1996). *Grammaticalization of the complex sentence: a case study in Chadic*. Amsterdam: John Benjamins.
- Greenberg J H (1963). *The languages of Africa*. Indiana University Research Center in Anthropology, Folklore, and Linguistics, publication 25. *International Journal of American Linguistics*, 29(1), part 2. Bloomington: Indiana University.
- Haruna A (2003). *A grammatical outline of Gùrdùṅ/Gùr-ùntùm (southern Bauchi, Nigeria)*. Westafrikanische Studien, Frankfurter Beiträge zur Sprach- und Kulturgeschichte, vol. 25. Cologne: Rüdiger Köppe.
- Hoffmann C (1963). *A grammar of the Margi language*. London: Oxford University Press.
- Jaggar P J (2005). 'Hausa and Chadic.' In Strazny P (ed.) *Encyclopedia of linguistics*. New York/London: Routledge. 445–447.
- Jaggar P J & Wolff H E (2002). *Chadic and Hausa linguistics: the selected papers of Paul Newman with commentaries*. Cologne: Rüdiger Köppe.
- Jungraithmayr H & Ibriszimow D (1994). *Chadic lexical roots, vol. 1: Tentative reconstruction, grading, distribution and comments; vol. 2: Documentation*. Sprache und Oralität in Afrika, 20. Berlin: Dietrich Reimer.
- Jungraithmayr H, Galadima N A & Kleinewillinghöfer U (1991). *A dictionary of the Tangale language (Kaltungo, northern Nigeria)*. Sprache und Oralität in Afrika, 12. Berlin: Dietrich Reimer.
- Newman P (2000). *The Hausa language: an encyclopedic reference grammar*. New Haven and London: Yale University Press.
- Newman P (2003). 'Chadic languages.' In Frawley W J (ed.) *International Encyclopedia of Linguistics*, 2nd ed., vol. 1. Oxford and New York: Oxford University Press. 304–307.
- Newman P & Ma R (1966). 'Comparative Chadic: phonology and lexicon.' *Journal of African Languages* 5, 218–251.
- Pawlak N (1994). *Syntactic markers in Chadic: a study on development of grammatical morphemes* [sic]. Warsaw: Instytut Orientalistyczny, Uniwersytetu Warszawskiego.
- Sachnine M (1982). *Dictionnaire lamé-français. Lexique français-lamé*. Paris: SELAF.
- Schuh R G (1981). *A dictionary of Ngizim*. University of California Publications in Linguistics 99. Berkeley and Los Angeles: University of California Press.
- Schuh R G (1998). *A grammar of Miya*. University of California Publications in Linguistics 130. Berkeley and Los Angeles: University of California Press.
- Stolbova O V (1996). *Studies in Chadic comparative phonology*. Moscow: Diaphragma Publishers.
- Wolff H E (1983). *A Grammar of the Lamang language (Gua'd Lamang)*. Afrikanistische Forschungen, 10. Glückstadt: J. J. Augustin.

Chibchan

A Constenla Umaña, University of Costa Rica, San José, Costa Rica

© 2006 Elsevier Ltd. All rights reserved.

The Chibchan stock is currently composed of the 16 languages from Central America and northwestern South America listed below with their main current alternate names, approximate number of speakers, and location: Pech (Paya; 900; Olancho Department, eastern Honduras), Rama (20; Rama Cay and other localities south of Río Escondido, southeastern Nicaragua), Maléku Jaíka (Guatuso; 300; Guatuso County, northern plains of Costa Rica), Cabécar (8500; Atlantic watershed and southern Pacific slope of the Talamanca Range, southern Costa Rica), Bribri (6000; southern Atlantic and Pacific

slopes of the Talamanca Range), Boruca (Brunka; 2, 20 semi-speakers with a passive domain of the language; Térraba Valley, southwestern Costa Rica), Teribe (a dialect of Naso; 3000; Teribe and Changuinola rivers area, northwestern Panama; Térraba, the Costa Rican dialect, is extinct), Buglere (Bocotá, Guaymí Sabanero; 3700; Bocas del Toro, Veraguas, Chiriquí Provinces, western Panama), Ngäbere (Guaymí; 110 000 in the Bocas del Toro, Chiriquí, and Veraguas provinces, Western Panama, and 2172 in the bordering area of southwestern Costa Rica), Kuna (70 000 in the eastern Atlantic coast and the southeastern Paya and Pucuro localities of Panama, and 800 in Arquía and Caimán Nuevo in the Urabá Gulf, Colombia), Chimila (450; lowlands to the south of Fundación River, Magdalena Department, Colombia), Cogui (Cágaba; 6000; northern, eastern,

and western slopes of the Sierra Nevada de Santa Marta, Colombia), Damana (Malayo; 1500; southern and eastern slopes of the Sierra Nevada de Santa Marta), Ica (Bintucua; 8000; southern slopes of the Sierra Nevada de Santa Marta), Barí (Motilón; 1500 in Colombia, 850 in Venezuela; Serranía de Motilones), and Tunebo (Uwa; 3500, mostly in Colombia, a few in Venezuela; eastern slopes of the Sierra Nevada de Cocuy). Formerly, the stock included at least eight more languages which are listed with their original location, and approximate time of extinction: Huetar (central Costa Rica, 18th century), Chánguena, Dorasque (both in western Panama, Chiriquí Lagoon area, beginning of the 20th century), Antioquian (central and northeastern Department of Antioquia, Colombia, 18th century), Tairona (the coast to the north of the Sierra Nevada de Santa Marta, 18th century or before), Kankuama (eastern slopes of the Sierra Nevada de Santa Marta, first half of the 20th century), Duit (Boyacá Department, Colombia, 18th century), and Muisca (Cundinamarca Department, Colombia, 18th century).

Subgrouping

The following subgrouping is based on both lexicostatistical and comparative evidence (Constenla, 1995: 42):

- I. Pech.
- II. Core Chibchan:
 - IIA. Votic: Rama, Guatuso.
 - IIB. Isthmic:
 - B1. Viceitic: Cabécar, Bribri.
 - B2. Boruca.
 - B3. Teribe.
 - B4. Guaymíic: Ngäbere, Buglere.
 - B5. Doracic: Dorasque, Chánguena.
 - B6. Kuna.
 - IIC. Magdalenic:
 - C1. Core Magdalenic:
 - C1.1. Southern Magdalenic:
 - C1.1a. Chibcha: Muisca, Duit.
 - C1.1b. Tunebo.
 - C1.2. Arhuacic:
 - C1.2a. Cogui.
 - C1.2b. Eastern-southern Arhuacic:
 - C1.2b.1. Eastern Arhuacic: Damana, Kankuama.
 - C1.2b.2. Ica.
 - C2. Chimila.
 - C3. Barí.

There are some indications that (a) the Isthmic group could be divided into two branches: Viceitic-Boruca and Teribe-Guaymíic-Doracic-Kuna, (b) the Magdalenic group could be also divided into two branches: Southern Magdalenic-Barí and Arhuacic-Chimila, (c) Huetar might belong to Votic, and (d) Tairona to Eastern-southern Arhuacic (Jackson, 1995: 67–68).

The split of Proto-Chibchan into the ancestors of Pech and Core Chibchan occurred, according to glottochronology, around 6550 years BP, at the times of the beginning of the transition from the hunter-gatherer way of life to the agricultural one. The greater diversity between the languages is found to the west and north, in Central America, which suggests that the Chibchan people's homeland must have been there, probably in Costa Rica and Panama, where archeology has found the oldest sites related to them.

External Relationships

There have been proposals of relationships between Chibchan and at least a score of other Amerindian language groups and isolates from Florida in the United States to northern Chile and Argentina (such as Timucua, Tarascan, Cuitlatec, Xincan, Lencan, Misumalpan, Chocoan, Andaquí, Beto, Warao, Yanomama, Paez, Barbacoan, Mochica, Kunza, Allentiac), which together would constitute a Macro-Chibchan phylum. None of these have been proved, and the quality of the supposed evidence in their favor is extremely poor (Constenla, 1993: 81–95).

Typology

The Chibchan languages belong to the Lower Central American Linguistic Area, characterized by features such as SOV order, postpositions, prepositional genitive, postpositive numerals and adjectives, lack of gender contrasts, and contrasts between voiced and voiceless stops.

The Chibchan languages of southern Costa Rica and western Panama, together with the Chocoan languages, constitute a Central Subarea characterized by the predominance of features such as distinctive vowel nasality, tense/lax vocalic contrasts, ergative or active case systems, and absence of person inflections. Most Chibchan languages in this subarea present numeral classifiers, postpositive demonstratives, and tone contrasts.

Pech, Rama, and Maléku Jaíka are part of a Northern Subarea, and the Magdalenic languages, of an Eastern Subarea. Although each of these subareas possesses its own characteristics, they share the predominance of features, both positive and negative,

opposed to those of the Central subarea such as accusative-nominative case systems (Maléku Jaika and Tunebo are exceptions to this), person inflection for possession in nouns and for agent and patient in verbs, prepositive demonstratives, and lack of numeral classifiers, distinctive vowel nasality, and tense/lax vocalic contrasts.

Bibliography

Constenla A (1991). *Las lenguas del Area Intermedia: introducción a su estudio areal*. San José: Editorial de la Universidad de Costa Rica.

Constenla A (1993). 'La familia chibcha.' In Rodríguez de Montes M L (ed.) *Estado actual de la clasificación de las lenguas indígenas de Colombia*. Bogotá: Instituto Caro y Cuervo. 75–125.

Constenla A (1995). 'Sobre el estudio de las lenguas chibchenses y su contribución al conocimiento del pasado de sus hablantes.' *Boletín Museo del Oro* 38–39, 13–55.

Holt D (1986). 'The development of the Paya sound system.' Ph.D. diss. University of California, Los Angeles.

Jackson R (1995). 'Fonología comparativa de los idiomas chibchas de la Sierra Nevada de Santa Marta.' *Boletín Museo del Oro* 38–39, 57–69.

Chimakuan Languages

E P Hamp, University of Chicago, Chicago, IL, USA

© 1994 Elsevier Ltd. All rights reserved.

The Chimakuan [tʃɪm'ækʊən] (or [ʔɪ]) family of languages on the northwest coast of North America comprises the smallest possible number of comparanda, just two known languages. This situation is not for lack of careful probing investigation, and the surrounding area is one of the best regions of aboriginal scholarship. The data will yield only with a fresh spurt of imagination.

The Family and its Recognition

This linguistic stock consists of Chemakum [tʃ'əməkəm], a neighbor's designation, and Quileute (also *Quillayute*, *Kwille'hiūt*, *Quilahutes*, *Kwe-dée-tut*) [kɥ'ili u-t] in English, /kʷo'li-yoʔ/ in their language, a tribe of less than 500 persons with 10 speakers in 1986, where we employ their self-name. The tribe occupies the western, Pacific coast of the Olympic Peninsula, state of Washington, USA, between the Wakashan Makah to the north and the Salishan Quinault to the south. Besides the Quileute at the river-mouth settlement of LaPush (a Chinook Jargon name, from French), the tribe includes the Hoh people /čālā-ʔ/ of the Hoh River (Quinault /hóx/); but note that in Quileute /čālā-l/ means 'Quinault language' – a tangle of important neighbors' designations, which looks like a language shift on the part of the Hoh.

Chemakum, now extinct, was located at a remove at the northeast corner of the Olympic Peninsula, adjacent to the Salishan Clallam, who absorbed them in 1890, and the Olympic mountains. In 1855

the Chemakum were about 100 strong, but by 1890, when F. Boas collected a modest number of words and sentences, there were only 3 speakers. The two language groups traditionally recognized their kinship, and Timofei Tarakanov reported this relation in his 1808–09 shipwreck account.

The remaining question of nearby kinship affects Salishan, which does not seem promising, or Wakashan, where the encouraging matches may be ancient borrowings.

The Data

The two languages share obvious characteristics; although some are of structural typology their cumulative weight coupled with phonological agreement in elements that make them up results in a strong case for homomorphism. The build of words is nearly identical. There are no prefixes, a couple of dozen inflexional suffixes, and over 200 elements called lexical suffixes, which additionally agree in selecting three empty bases with idiosyncratic semantics. Diminutives and plurals use infixes. There are identical structure requirements in word classes: for every predicate, for article and pronoun suffixes, and for (non-) feminization of deictics. More of an areal feature here is the property that all words except particles can be predicates.

There are phonological regularities and morphological divergences that lead to the recognition of correspondences with time depth. Ironically, such non identities are necessary to demonstrate kin relation. The consonants agree well, the fullest set is the glottalized one: *pʰtʰččkʰwqʰqʰwʰ, plus *mʰnlyw matching *mnlyw. A set of spirants matching *ʎ through ʔ fills out a picture that in its outer limits

points more to areal typology. An interesting idiosyncrasy is the lack of plain lateral λ , which Quileute has filled in, this writer believes by a Grassmann-like dissimilation. Just as in pre-Chimakuan before front vowel labio-velars palatalized to palatal groove obstruent, so did front velars in Chemakum. Quileute has developed vowel length and pitch accent (or, as this writer believes, more a stress placement) and has undergone stress shift to penult.

Quileute, probably the only language in the world to lack surface nasals completely, has turned them into voiced stops; but the witch Daski-ya of folklore spoke in her characterizing style with nasals.

These are powerful correspondences. Chemakum seems to have revalued its plural on the Clallam model. Quileute may have lost detail in the subject pronouns, and perhaps mirrors Tillamook in the feature-inflexion of feminines.

Bibliography

- Andrade M J (1931). *Quileute Texts*. Columbia University Contributions to Anthropology 12, New York [reelited tapes by E P Hamp in Americal Philosophical Society Library].
- Andrade M J (1933). Quileute. In Boas F (ed.) *Handbook of American Indian Languages*, part 3. New York.
- Andrade M J (1953). Notes on the relations between Chemakum and Quileute. *IJAL* 19, 212–215.
- Jacobsen W H Jr (1979). Chimakuan comparative studies. In Campbell L & Mithun M (eds.) *The Languages of Native America*. Austin, TX: University of Texas.
- Suttles W (ed.) (1990). *Northwest Coast*, vol. 7 of *Handbook of North American Indians*. Washington, DC: Smithsonian Institution.

Chinantec: Phonology

D Silverman, Nanvet, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Chinantecan is a group of about 14 VSO languages within the Otomanguean family, spoken by approximately 90 000 people in northeastern Oaxaca, Mexico, having branched from the Otomanguean tree more than 16 centuries ago. The 14 major languages (where ‘language’ is defined as a speech community with mutual intelligibility not in excess of 80% with other communities) are Ojitlán, Usila, Tlacoatzintepec, Chiltepec, Sochiapan, Tepetotutla, Tlatopusco, Palantla, Valle Nacional, Ozumacín, Lalana, Lealao, Quiotepec, and Comaltepec. The first seven are northern languages and tend to be more innovative phonologically; the second seven southern languages are more conservative. Syllables are usually CV, with only a few post-vocalic elements, among them a nasal and/or laryngeals. Proto-Chinantec is reconstructed as possessing consonants $*p$, $*t$, $*k$, $*k^w$, $*b$, $*z$, $*g$, $*g^w$, $*s$, $*m$, $*n$, $*\eta$, $*w$, $*l$, $*r$, and $*j$. Laryngeals $*h$ and $*\text{ʔ}$ could stand alone pre-vocally, or could precede any of the voiced consonants. Additional consonant-glide clusters are reconstructed as well. The reconstructed tonal inventory includes $*H$, $*L$, $*HL$, $*LH$, and $*HLH$. Vowels included $*i$, $*e$, $*a$, $*u$, $*i$, and $*a$, as well as several diphthongs. The vowels may be augmented in a bewildering number of ways, however. In modern Comaltepec – the most conservative Chinantecan

language – eight vowel qualities (i , e , æ , a , o , ʌ , i , u) may be combined with five tonal qualities (L , M , H , LM , LH), two voice qualities (plain and aspirated), a nasality contrast, as well as a binary length contrast. The cross-classification of these 5 independent systems results in 320 possible nucleus qualities ($8 \times 5 \times 2 \times 2 \times 2$). Thus, a single vowel quality may possess up to 40 contrastive values.

Chinantec roots and words are usually monosyllabic. The rich inflectional system normally involves modification of root vowels, resulting in monosyllabic stems that bear a particularly high informational load. In Comaltepec, for example, a single syllable may contain not only the root but also (in the case of verb complexes) active/stative markers, gender markers (animate/inanimate), transitivity markers (intransitive/transitive/ditransitive), aspect (progressive/intentive/completive), and possibly subject pronoun clitics (two subsyllabic classes). Methods of stem modification involve nasalization, tone, length, phonation augmentation, and sometimes consonant changes. Additionally, certain irregular patterns are marked by ablaut. Due to their inherent inflection, bare verbal roots do not exist as such in Chinantecan. All Chinantecan languages have a large number of verb classes, along with many lexical exceptions. Classes are differentiated by patterns of identity or nonidentity across aspect/person combinations. For example, in the partial paradigm for the verb ‘to hit’ shown in Table 1, some complexes are identical to others, while others are different. Verbs in this class

Table 1 Partial verb paradigm from comaltepec

| | | | | |
|----------------------------|-------|-------|-------|-------|
| hit (transitive/inanimate) | 1s | 1p | 2 | 3 |
| progressive | bah↓ | ba↓ | bah↓ | bah↓ |
| intensive | bah↓ | bah↓ | bah↓ | bah↓ |
| completive | bah↓ | bah↓ | bah↓ | bah↓ |
| hit (transitive/animate) | | | | |
| progressive | bA:ꞑ↓ | bA:ꞑ↓ | bA:ꞑ↓ | bA:ꞑ↓ |
| intensive | bA:ꞑ↓ | bA:ꞑ↓ | bA:ꞑ↓ | bA:ꞑ↓ |
| completive | bA:ꞑ↓ | bA:ꞑ↓ | bA:ꞑ↓ | bA:ꞑ↓ |

Table 2 Examples of stem inflection in Quiotepec (Robbins, 1968)

| | |
|------------------------|--|
| k ^w o:↓ | I give (something) |
| k ^w o:ꞑ↓ | I gave (something) |
| k ^w o:ꞑꞑ↓ | thou givest (something) |
| k ^w o:ꞑꞑ↓ | thou gavest (something) |
| k ^w o:ꞑꞑ↓ | I give (something to someone) |
| k ^w o:ꞑꞑ↓ | I gave (something to someone) |
| k ^w o:ꞑꞑ↓ | thou givest (something to someone) |
| k ^w o:ꞑꞑ↓ | thou gavest (something to someone) |
| k ^w o:ꞑꞑꞑ↓ | I give (something animate) |
| k ^w o:ꞑꞑꞑ↓ | I gave (something animate) |
| k ^w o:ꞑꞑꞑ↓ | thou givest (something animate) |
| k ^w o:ꞑꞑꞑ↓ | thou gavest (something animate) |
| k ^w o:ꞑꞑꞑꞑ↓ | I give (something animate to someone) |
| k ^w o:ꞑꞑꞑꞑ↓ | I gave (something animate to someone) |
| k ^w o:ꞑꞑꞑꞑ↓ | thou givest, gavest (something animate to someone) |

will tend to show a similar pattern of identity and nonidentity across cells, while verbs in other classes show a different pattern.

Table 2 provides examples of stem inflection from Quiotepec (Robbins, 1968).

In at least some Chinantecan languages, the verb may be prefixed by a subject agreement marker for intransitive verbs, or by an object agreement marker for transitive verbs. Additional verbal prefixes include a negation marker, and tense and aspect markers (imperfect, past, hodiernal past, perfect, past imperfect, etc.). Unlike verbs, nouns do not typically display internal inflection, instead showing stability across inflectional augmentation. In Tepetotutla, for example, noun roots may concatenate with a quantifier, a gender-inflected numeral, a classifier, etc. In Lealao, constituents of the noun phrase may include a quantifier, the head, a modifier, a possessor, and a deictic marker, in that order, as well as a classifier prefix in some cases.

Stem complexes are obligatorily stressed. Posttonic and pretonic syllables are not stressed. Stressed syllables may possess greater phonological and morphological complexity than do unstressed syllables. In Sochiapan, unstressed syllables differ from stressed

ones in displaying a more limited distribution of phonemes. Posttonic syllables in Palantla consist of a small list of words that do not contrast for tonal features. Pretonic syllables, while maintaining tonal contrasts, do not possess postvocalic elements, except in very careful speech. In Comaltepec, posttonic syllables consist of a limited set of clitics, person-of-subject inflectors (in verbs), and possessors (in nouns). Pretonic syllables consist of only several verbal prefixes and a few proclitics, and possess a smaller inventory of tone values. These syllables are not a site for further inflection, and thus do not possess morphological complexity. In Quiotepec, too, stress falls on the major lexical classes (verbs, nouns, etc.); most pretonic syllables consist of inflectional material. Pretonic syllables only occur with single tones, never with tonal contours. In at least several Chinantecan languages, the vocalism of posttonic syllables is harmonically determined by the stem vowel. Tone may spread from stem to suffix as well.

Regarding Chinantecan stress, several languages are traditionally characterized as possessing either ‘ballistic’ stress or ‘controlled’ stress on stem syllables. In Palantla, Tepetotutla, Sochiapan, and Comaltepec, ballistic syllables have been characterized by an initial surge and rapid decay of intensity, and a loss of voicing of postvocalic elements; controlled syllables exhibit no such initial surge of intensity, displaying a more evenly controlled decrease of intensity, and a lack of postvocalic devoicing. Ballistic syllables tend to be shorter in duration than controlled syllables, and may possess a smaller inventory of tonal patterns. In at least several Chinantecan languages, ballistic syllables cross-classify with almost every other syllable type. Both oral and nasal vowels, both long and short vowels, preaspirated and preglottalized onsets and plain onsets, open and checked syllables, and nasally closed syllables, may all possess ballistic stress. Ballistic stress interacts most significantly with tone, tending to raise high tones and lower low tones. In Lalana, ballistic stress (considered postvocalic h in some analyses) may not occur with glottal checking, and may occur with only H, L, and HL tones, whereas controlled syllables reportedly also possess MH, LH, and HLH, and may be checked. In Lealao, only level tones (L, M, H, VH) may occur with ballistic stress, whereas controlled syllables may also occur with tonal contours (LM, LH). In Comaltepec, ballistic syllables may occur with almost any tonal pattern.

The ballistic stress found in some Chinantec languages corresponds to tonal lowering in Ojitlán and Usila. Quiotepec is variously characterized as possessing ballistic accent or raised tones in these same contexts, often accompanied by postvocalic aspiration. The Chinantecan ballistic syllable

Table 3 Tone sandhi in Comaltepec

| Non-sandhi context | Sandhi context | Gloss |
|--------------------|------------------------|------------------|
| to:ɿ | kwaɿ to:ɿ | give a banana |
| ɿhɿ | kwaɿ ɿhɿ | give a chayote |
| kuz:ɿ | kwaɿ ku:ɿ | give money |
| hi:ɿ | mi:ɿ hi [○] | I ask for a book |
| mohʔ:ɿ | mi:ɿ mohʔ [○] | I ask for squash |

corresponds to postvocalic aspiration in related Mixtecan and Otopamean languages, to prevocalic aspiration in related Popolocan languages, and to glottally ‘interrupted’ (CVʔV) syllables in the Chatino, Zapotec, and Tlapanec languages. Chinantecan ballistic syllables may derive from Proto-Otomanguean *CVh syllables (which may or may not have been phonetically realized as interrupted vowels). Indeed, recent phonetic and phonological investigations have recharacterized the ballistic phenomenon as largely laryngeally-based, involving postvocalic aspiration.

Segmental sandhi is rather limited in Chinantecan, although tone sandhi is widespread, being both phonologically and morphologically conditioned. The best-studied tone sandhi system is that of Comaltepec. Here, LH tones spread their H component on to a following vowel. Furthermore, M tones on unchecked controlled syllables (deriving from Proto-Chinantec H) trigger the presence of an H tone on the following syllable. Examples are shown in Table 3.

Bibliography

- Anderson J L (ed.) (1989). *Comaltepec Chinantec syntax. Studies in Chinantec languages* (vol. 3). Dallas: Summer Institute of Linguistics.
- Merrifield W R (1968). *Palantla Chinantec Grammar. Papeles de la Chinantla V. Serie Científica 9*. Mexico: Museo Nacional de Antropología.
- Merrifield W R & Rensch C R (eds.) (1990). *Syllables, tone, and verb paradigms. Studies in Chinantec languages* (vol. 4). Dallas: Summer Institute of Linguistics.
- Rensch C R (1968). *Proto Chinantec phonology. Papeles de la Chinantla VI. Serie Científica 10*. Mexico: Museo Nacional de Antropología.
- Rensch C R (1976). *Comparative Otomanguean phonology*. Indiana University: Bloomington.
- Rensch C R (ed.) (1989). *An etymological dictionary of the Chinantec languages. Studies in Chinantec languages* (vol. 1). Dallas: Summer Institute of Linguistics.
- Robbins F E (1968). *Quiotepec Chinantec grammar. Papeles de la Chinantla IV. Serie Científica 8*. Mexico: Museo Nacional de Antropología.
- Rupp J E (ed.) (1989). *Lealao Chinantec syntax Studies in Chinantec languages* (vol. 2). Dallas: Summer Institute of Linguistics.
- Silverman D (1997). ‘Tone sandhi in Comaltepec Chinantec.’ *Language* 73, 473–492.
- Silverman D (1997). *Phasing and recoverability*. New York: Garland.
- Westley D O (1991). *Tepetotutla Chinantec syntax Studies in Chinantec languages* (vol. 5). Dallas: Summer Institute of Linguistics.

Chinese

Y Gu, Chinese Academy of Social Sciences, Beijing, China

© 2006 Elsevier Ltd. All rights reserved.

The State of the Art

If language is ultimately seated in the minds of individual speakers, as some linguists claim, then Chinese can be described as a collection of over 1.3 billion idiolects scattered around the world, in Mainland China, Taiwan, Hong Kong, and Singapore in particular. If on the other hand language is held to be the property of a speech community, as many linguists believe, Chinese is then an assemblage of numerous ‘dialects’ spreading over different continents and across time zones, some of which are so

different that their speakers cannot even communicate with one another. In spite of the vast diversity, and even some mutual oral unintelligibility, all literate speakers can overcome the barrier imposed by the oral unintelligibility via reading (not aloud!) and writing. The writing script partly enables the users to transcend the differences of idiolects and dialects, and bridges the past and the present.

In this article, Chinese will be discussed within its two natural divisions: spoken Chinese and written Chinese. The former includes (1) the classification of dialects and their geographic and demographic distributions; (2) Putonghua as a lingua franca; and (3) a brief discussion plus sound illustrations of three major dialects. The latter includes (1) the writing script, and (2) the historical evolution of written Chinese from archaic Chinese to modern Chinese.

The article concludes with a summative account of how Chinese, both spoken and written, is electronically processed.

Spoken Chinese

Although Chinese, like any other language in the world, is substantiated in idiolects, i.e., parole in the Saussurean term, they are thrown away after being used as evidence for language system construction i.e., the Chinese language. In other words, talking about Chinese, over 1.3 billion idiolects are generally ignored. What linguists are interested in is the various dialects evolved from them. The number of dialects depends on how fine-grained the researcher's scheme is intended to be. It is hardly a rare case that people in two villages only a dozen of miles apart cannot intelligibly communicate through speech.

Dialect Classification and Distribution

Chinese dialects can be classified by adopting a tree-like structure. The first branching-out from the trunk is the two major supergroups: Mandarin and non-Mandarin. Mandarin includes eight subgroups: Northeastern, Beijing, Beifang, Jiaoliao, Zhongyuan, Lanyin, Southwestern, and Jianghuai. The non-Mandarin group comprises nine subgroups: Jin, Wu, Hui, Gan, Xiang, Min, Yue, Pinghua, and Hakka. Each of the subgroups has its own clusters, each of which in turn encompasses local dialects (see Figure 1).

Geographically speaking, Mandarin is spoken in the following provinces and major cities: Heilongjiang, Jilin, Liaoning, the eastern part of the Inner Mongolia Autonomous Region, Shandong, Beijing, Tianjing, Hebei, Shanxi, Gansu, Qinghai, Ningxia, Hui Autonomous Region, Sichuan, Yunnan, Guizhou,

Guangxi Zhuang Autonomous Region, western part of Hubei, Chongqing, northern parts of Jiangsu, and Anhui. The total Mandarin-speaking population, based on the 1982 census, was about 662.23 million. Table 1 shows the demographic distributions among the subgroups of Mandarin. The demographic distributions of other non-Mandarin dialects are shown in Table 2.

Table 1 Mandarin-speaking population by 1982

| | |
|-------------------|------------------|
| Northeastern | 82.00 |
| Beijing | 18.02 |
| Beifang | 83.63 |
| Jiaoliao | 28.83 |
| Zhongyuan | 169.41 |
| Lanyin | 11.73 |
| Southwestern | 200.00 |
| Jianghuai | 67.25 |
| Yet to be grouped | 1.36 |
| Total | 662.23 (million) |

Table 2 Demographic distributions of other non-Mandarin dialects by 1982

| | |
|-------------------|------------------|
| Jin | 45.70 |
| Wu | 69.75 |
| Hui | 3.12 |
| Gan | 31.27 |
| Xiang | 30.85 |
| Min | 55.07 |
| Yue | 40.21 |
| Pinghua | 2.00 |
| Hakka | 35.00 |
| Yet to be grouped | 2.06 |
| Total | 315.03 (million) |

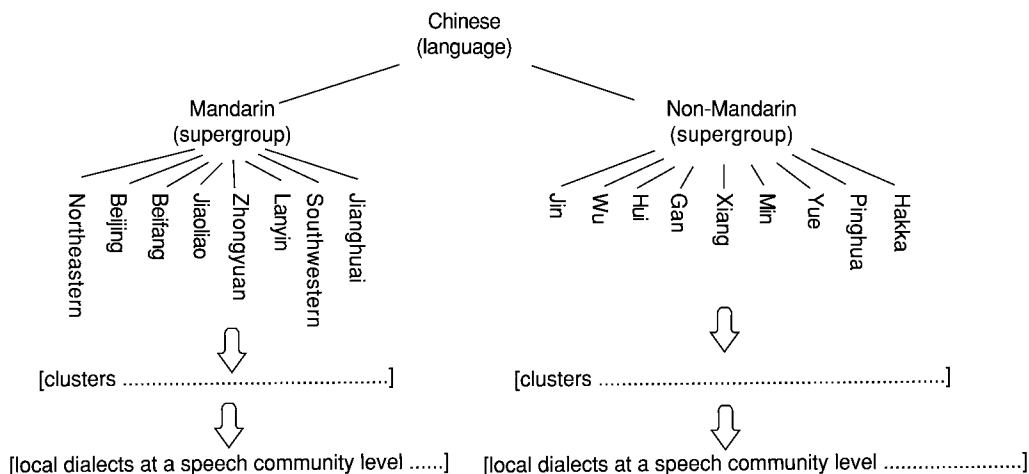


Figure 1 Classification of Chinese dialects.

Mandarin Chinese is often nontechnically regarded as an equivalent to Chinese, which was historically the language of the Han nationality. Thanks to massive immigration and frequent contact, Mandarin Chinese is spoken by non-Han ethnic peoples as well. Some members of the Hui nationality, for instance, who are of Mohammedan origin, adopt Mandarin as their mother tongue. Almost all members of the She and Manchu nationalities speak Mandarin Chinese. Conversely, some people of Han origin in Hainan Province speak the Be language instead of Mandarin.

Putonghua as Lingua Franca

Dialects create diversity and local identity, and at the same time impose constraints on communication and social interaction. A tension always exists between diversification and standardization of the language. Many campaigns have been launched in the long history of China in favor of standardizing both spoken and written Chinese. The policy of *shu tong wen zi* ('writing according to the same script') adopted in the Qin Dynasty (248–207 B.C.) was in fact a systematic reform undertaken by the imperial court to standardize the writing script. In the Sui Dynasty, Lu Fayan's (fl. 600 A.D.) *Qieyun* ('Guide to poetic rhyming') became a standard reference on pronunciation for the generations to come, as well as for the reconstruction of ancient phonological systems. The campaign for the standardization of modern Chinese started as early as the last leg of the Qing Dynasty (1616–1911 A.D.) when the National Language Movement was vigorously launched as a part of the measures to revitalize the shattered country. It was argued that the nation could not be unified without a unified language. Guoyu ('national language') was initially envisaged and artificially constructed on the basis of some major dialects. This proved to be untenable, for it was next to impossible to promote such a language without natural speakers. New Guoyu ('new national language'), with the Beijing dialect as its base, was proposed and eventually adopted. Immediately after the founding of the People's

Republic of China in 1949, language reform was put high on the government's agenda. Modern Standard Chinese, officially called Putonghua, was adopted as the national language. It uses the Beijing dialect for its standard pronunciation and northern dialects as its base input. Putonghua is officially stipulated to be the language of instruction at all levels of education, and of mass media.

The term Guoyu is still being used in Taiwan, while in Singapore it is called Huayu (i.e., Chinese). Putonghua, Guoyu, and Huayu are three different terms to refer to more or less the same Modern Standard Chinese.

Modern Standard Spoken Chinese

Phonology The phonological structure of Modern Standard Chinese is conceptualized more in traditional Chinese terms than otherwise. A syllabic structure has three essential components: initials, finals, and tones. The initials and finals are two segments of a syllable, while the tones are supersegmental, i.e., features superimposed on the segments. The initials are the sounds known as consonants in Western literature. The finals, i.e., vowels, have internal structures of their own: the medial and the root of the final, which is further decomposed into two: the main vowel and the syllabic terminal (see Figure 2).

The initial, the medial, and the syllabic terminal are not obligatory to make a Chinese syllable. A simple syllable can consist of a main vowel plus a tone only.

The possible initials, finals, and tones of Modern Standard Chinese are summarized in Tables 3, 4, and 5, respectively.

It is perhaps well-known now to the non-Chinese speaking world that Chinese tones are phonemic, that is, the same phonetic syllable pronounced in different tones will produce different words. The syllable /ma/ is the classic example: ma55 (mother), ma35 (hemp), ma214 (horse), ma51 (scold), and ma0, (a functional particle without a fixed lexical meaning).

While tones are properties of words, there are also intonations of utterances. The relation between the tone and the intonation is often metaphorized as

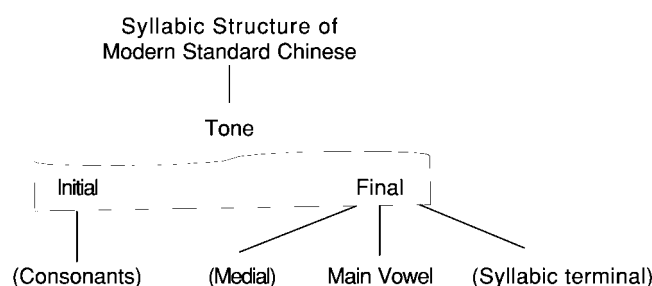


Figure 2 Syllabic structure of Modern Standard Chinese.

small ripples (cf. word tones) riding on large waves (cf. utterance intonations). The interaction between the tone and the intonation results in an algebraic sum of the two kinds of waves.

Grammar It is generally held that, although Chinese dialects are so diversified that mutual unintelligibility in speech is not uncommon, they are conversely amazingly unified in matters of grammar. There are some minor divergencies found between dialects, for example, with regard to the order of direct and indirect objects, the Wu dialects and Cantonese differing from Mandarin Chinese. Cases like this, however, are

extremely limited. It is quite valid to hold that there is one universal Chinese grammar.

At the risk of oversimplification, which is unavoidable in such a short essay as the present one, Chinese grammar, in comparison with English and other European languages, is pragmatically oriented. The subject and predicate in the grammar of Western languages are best viewed as the topic and comment in Chinese. The subject/actor and the predicate/action are treated as a special case of topic and comment. For instance, *jiu bu he, yan chou* (word-for-word rendering: wine not drink, cigarette smoke) is understood as 'Talking about wine, I don't drink; but as for cigarettes, I do smoke.'

The topic-comment structure has something to do with the complaint often made by Westerners about Chinese saying 'no' but actually meaning 'yes.' Responding to the utterance *zhe shu bu hao* (word-for-word rendering: 'this book not good'), if the speaker also thinks that the book is not good, he will say *shi* ('yes'), meaning that he agrees with what the first speaker said about the book. While the English mind checks the statement against the fact, the Chinese mind expresses agreement or disagreement with the speaker. In other words, the Chinese mind tends to treat the speaker's utterance as setting up a topic, and the responder's job is to comment on the topic. The issue of the truth or falsehood of the statement becomes secondary.

Pragmatics One of the Chinese politeness maxims dictates that the speaker should denigrate him or herself, while elevating the other. This maxim has been codified in a range of lexical items. All the self-related expressions, including those referring to one's family members, relatives, properties, writings, and so on, are marked with denigration, whereas the

Table 3 Initials of modern standard Chinese

| Description | Pinyin | IPA |
|------------------|--------|-----|
| Bilabials | b | p |
| | p | p' |
| | m | m |
| | f | f |
| Alveolars | d | t |
| | t | t' |
| | n | n |
| | l | l |
| Dental sibilants | z | ts |
| | c | ts' |
| | s | s |
| Retroflexes | zh | tʂ |
| | ch | tʂ' |
| | sh | ʂ |
| | r | ɻ |
| Palatals | j | tɕ |
| | q | tɕ' |
| | x | ç |
| Velars | g | k |
| | k | k' |
| | h | x |

Table 4 Finals of modern standard Chinese

| Pinyin | IPA | Pinyin | IPA | Pinyin | IPA | Pinyin | IPA |
|--------|-----|--------|-----|--------|-----|--------|-----|
| | | i | i | u | u | ü | y |
| a | a | ia | ia | ua | ua | | |
| o | o | | | | uo | | |
| e | ɤ | | | | | | |
| ê | ɛ | ie | iɛ | | | üe | ye |
| ai | ai | | | uai | uai | | |
| ei | ei | | | uei | uei | | |
| ao | au | iao | iau | | | | |
| ou | ou | iou | iou | | | | |
| an | an | ian | ien | uan | uan | üan | yen |
| en | ən | in | in | uen | un | ün | yn |
| ang | aŋ | ing | iŋ | ueng | uəŋ | | |
| ong | uŋ | iong | ɨŋ | | | | |
| er | ər | | | | | | |

other-referring expressions, including those referring to the other's family members, relatives, properties, writings, and so on, carry the force of elevation. For instance, a man referring to his own house will politely use *han she* ('cold living place'), but *fushang* ('mansion') to refer to the other's residence.

The self-denigration and other-elevation maxim also operates in compliment-taking. Complaints are made about Chinese failing to take a compliment gracefully. Hearing a compliment like *ni de yifu hen piaoliang* ('your dress is beautiful'), a Chinese lady will vigorously insist that it is very ugly indeed: *bu bu bu, chou shile* ('no no no, deadly ugly').

Written Chinese

In most languages, spoken and written forms are generally regarded as two functional varieties of one and the same language. The relation between spoken and written Chinese, however, cannot be dealt with so readily in the same way. Inscriptions incised on oracle bones, dated 1400–1100 B.C., are the earliest existent written records of Chinese. The inscriptions were not transcripts of the speeches of emperors or tribal kings. They can be regarded, at best, as setting up some of the earliest instances of a particular genre of written Chinese. By the time of the late years of Qing Dynasty (1616–1911) 3000 years or so later, the archaic written Chinese had become so different from the contemporary spoken Chinese that it would take years of dedicated study before one could

read and write it. To make things even worse, the archaic written Chinese was prescribed as the medium of education. It was, and still is, no easier for students to learn it than it would be to learn a foreign tongue. Some language reform activists in the 1910s went on record arguing that archaic written Chinese was partially to blame for the humiliating decline of the Chinese civilization following the time of the Tang dynasty (618–907).

Attempts to reform written Chinese thus had two aspects: the reform of the writing script and the reform of archaic written Chinese as the medium of education.

Writing Script Reform: Alphabetization Versus Simplification

The nature of the Chinese writing script has been disputed for years, as can be seen in the variety of English terms used to designate the marks on paper known in Chinese as *hanzi* (i.e., Chinese characters): pictographs, pictograms, ideographs, ideograms, phonograms, logographs, ideophonographs, lexigraphs, morphographs, sinographs, and so on. The evidence for the claim that the Chinese writing originated from picture-drawing is substantial. Table 6 shows four instances of pictographs taken from oracle bone inscriptions with their corresponding present-day characters.

It is apparent that the pictographs have evolved, through orthographic reforms, to such an extent that even those characters with highly iconicized origins as shown in the table have lost their picturesqueness. Chinese characters are constructed from five basic strokes (see Table 7) in a square space.

Picture-based character creation is only one of the many ways in which Chinese characters are constructed. Some philologists in the Han dynasty (206 B.C.–220 A.D.), on the basis of the then existent writings, abstracted six principles of character formation. Later studies show that only four of them are genuine: (1) *zhi shi*, the simple indicative principle; (2) *xiang xing*, the pictographic principle; (3) *hui yi*, the compound indicative principle; and

Table 5 Tones of modern standard Chinese

| Chinese terms in Pinyin | Description | Value in five-point scale |
|-----------------------------------|----------------|---------------------------|
| Yinping (1 st tone) | high level | 55 |
| Yangping (2 nd tone) | rising | 35 |
| Shangsheng (3 rd tone) | falling-rising | 214 |
| Qusheng (4 th tone) | falling | 51 |
| Qingsheng (neutral tone) | | 0 |

Table 6 Instances of Chinese pictographs

Pictographs found in oracle bone inscriptions

Corresponding present-day characters

English translation



虎

tiger



鹿

deer



馬

horse



象

elephant

Table 7 Strokes and character writing

| Strokes | Sample | Translation |
|---------|--------|-------------|
| 一 | 舞 | dance |
| 一 | | |
| ㇏ | | |
| ㇏ | | |
| ㇏ | | |

(4) xing sheng, the semantic–phonetic principle. The pictographic method of character formation had ceased to be productive by the Han dynasty. The semantic–phonetic character formation has been the most productive of all, and the majority of Chinese characters are thus constructed. It is on this account that the Chinese writing system can be appropriately designated as being morphosyllabic.

A Chinese character can be as simple as one stroke (e.g., 一 ‘one’), or as complex as dozens of strokes (e.g., 𪗇 ‘snuffling’). Given a set of 11 834 characters, the average number of strokes per character is 11.5516, and 63 percent of the set is made of 12-stroke characters. Since it is quite a challenging task to learn to write such characters, there has been no shortage of appeals to reform the writing script. As early as the 1910s, some language reform activists argued for abolishing the characters altogether, to be replaced with a new alphabet script. This proved to be completely infeasible. The PRC government eventually adopted three reform measures: (1) a romanization alphabet known as Pinyin that is used to mark the pronunciations of characters; (2) a simplification scheme according to which 1754 characters would be simplified; and (3) a total of 1055 duplicate characters that were to be abolished.

The Reform of Archaic Written Chinese

Archaic written Chinese models the writings prevalent from the Spring and Autumn (770–476 B.C.) to the Later Han (25–220 A.D.) periods. Partially because the characters were immune to the dynamic changes of actual speech sounds over space and time, archaic written Chinese achieved, as it were, an independent symbolic existence. By the 1900s, it had no natural speakers. It did, however, have several potential rivals under the name *baihua wen*, literally meaning ‘unadorned speech writing,’ which was far closer to the contemporary vernacular speech. The reform movement basically dethroned archaic

written Chinese and replaced it with the *baihua wen* that had been formerly much despised. The reform proved to be an uphill task, however, as it met with fierce resistance from die-hard adherents.

Three Major Dialects

As graphically shown in Figure 1, the non-Mandarin supergroup falls into nine subgroups of dialects, which of course can be further divided into smaller groups. In this section three dialects, Hong Kong Cantonese, Shanghainese and Fuzhou dialect, representing the Yue group, the Wu group and the Min group respectively, are examined as a window to show what the non-Mandarin dialects look like. They are highlighted here thanks to the demographic size (see Table 2) and relatively prestigious status they enjoy.

The Yue Group: Hong Kong Cantonese

Hong Kong Cantonese is one of the important varieties of the Yue group. It is spoken by 89 percent of Hong Kong’s 6.4 million population (by the 1996 census) in family discourse. It is also used in some radio and TV programs, and as an instructional language in schools and university classrooms. English was the main official language in the former British colony, but its use actually was, and still is, quite limited. Since the return of sovereignty to China in 1997, Putonghua has become increasingly popular. Having said this, Hong Kong Cantonese still remains a true vernacular of the local people.

The term ‘Cantonese’ is derived from Guangzhou, the most influential city in southern China, which is known as Canton in English. Hong Kong and Guangzhou Cantonese are not noticeably different except that the former’s lexicon has more English loan words than the latter’s. In speech Cantonese and Mandarin or Putonghua are mutually unintelligible. Educated Cantonese speakers, however, use the standard form of written Putonghua. There are some spoken Cantonese words that have no corresponding Putonghua characters. Some Cantonese written words coined by local newspapers and in advertisements in Hong Kong are unintelligible to Putonghua readers.

Backed up by the economic and financial strength and influence of Hong Kong and Guangzhou, Cantonese is enjoying a prestige that is unprecedented for any regional dialect in China, and is the most studied of all the dialects. Grammars, dictionaries, and textbooks have been written to render it more like a language than a regional dialect.

Cantonese has 16 initial consonants. Unlike Mandarin, it has completely nasal syllables with *m* and *ng* functioning as vowels. For instance, the Cantonese word for the Mandarin word *wu* ('five') is *ng*, which can only be a syllabic nasal terminal of a final in Mandarin. It has eight vowels, and two sets of consonants that can be syllabic terminals: (1) nasals: -*m*, -*n*, -*ng*; (2) unreleased consonants: -*p*, -*t*, -*k*. Its tone system is far more complex than that of Putonghua. The exact number of tones is not without controversy. Some hold that only six tones are clearly distinctive in Hong Kong Cantonese, although there can be up to nine tones in the Yue group.

The Wu Group: Shanghainese

The Wu group is spoken mainly in Shanghai, Southern Jiangsu Province, and a large part of Zhejiang Province. Historically the Suzhou Wu dialect enjoyed more prestige and esteem than the other regional varieties. When Shanghai established itself as an industrial and commercial center in China, it lost its glory and was replaced by Shanghainese, whose speakers seem to be eager to establish their own identity. Shanghainese speakers, who may speak fluent Putonghua, will lose no opportunity to code-switch to Shanghainese if they can be understood by an interlocutor, even at the risk of rudely shutting off any non-Shanghainese speakers from the conversation.

In comparison with Cantonese, Shanghainese is very much under-studied. Existent literature on it mainly consists of academic research papers. Like Cantonese, educated Shanghainese speakers write in written Putonghua, although there exist lexical items that are unique to the dialect.

The term Shanghainese refers to the majority speech of downtown Shanghai. It has 28 initials (i.e., consonants), and 43 finals (i.e., vowels). One of its hallmark features (and also of the Wu group) is a three-way distinction in the initial consonants *p*, *p'*, and *b*, which become a two way distinction, *p*, and *p'*, in Putonghua. Although Wu dialects have seven or eight tones, tones 4, 5, and 6 have been lost as separate categories, which results in five tones in Shanghainese: (1) high level (53), (2) level high (35), (3) low level (13), (4) high + a glottal stop (5), and (5) low + a glottal stop (1).

The Min Group: Fuzhou Dialect

The Min group is mainly spoken in Fujian, Taiwan, Hainan, as well as some areas in Guangdong, Zhejiang, Guangxi, and Jiangxi. It is by no means a

Table 8 Sample usage in Fujian dialect

| Putonghua | Fujian dialect | English translation |
|--------------|----------------|---------------------|
| 水稻 (shuidao) | 糶 | rice |
| 书信 (shuxin) | 批 | letter |
| 冷 (leng) | 清 | cold |
| 哭 (ku) | 啼 | cry |
| 逃跑 (taopao) | 走 | escape |
| 走 (zou) | 行 | walk |

homogeneous group. On the contrary, even within Fujian Province six subgroups can be identified, one of which is known as the Min eastern subgroup, with the Fuzhou dialect as its prototype. Mutual communicability within this eastern subgroup is quite low.

Historically the Fuzhou dialect was understood to cover an area of 11 counties. The present-day use of the term is much restricted to the speech of the locals in downtown Fuzhou. Phonologically it has 15 consonants, 46 vowels including diphthongs, and 7 tones. One of the striking features of the Fuzhou dialect in comparison with Mandarin is that it has preserved a great many archaic words or usages. For instance, the word for 'rice' is 糶 in Fujian dialect, which is totally obsolete in Putonghua. For another instance, the word 批 in Fujian dialect is used to mean a letter, a usage found only in archaic Chinese. Table 8 lists some more instances.

Sound Illustrations

The phonological differences between Putonghua, Hong Kong Cantonese, Shanghainese, and Fujian dialects can be illustrated by the ways four natural objects – the sun, the moon, stars, and thunder – are lexicalized and pronounced (see Table 9).

Chinese Information Processing

At the early stage of computer technology, processing Chinese characters seemed to be such a forbidding task that calls for the romanization of the Chinese writing system were made again, but initial conceptions of the problem proved to be exaggerated. The national standard GB 2312–80, established on the basis of ISO 646 and officially coming into effect in 1981, provides a standard scheme of coding 6763 characters, which are subdivided into two groups according to the frequency of usage: the most frequent set, and the less frequent set. The most frequent set of 3755 characters is assumed to be 99.9% adequate for general usage (based on a statistical study of lexical frequency made in 1974). The GB

Table 9 The phonological differences between Putonghua, Hong Kong Cantonese, Shanghainese, and Fujian dialects

| <i>Dialects</i> | <i>The sun</i> | <i>The moon</i> | <i>Stars</i> | <i>Thunder</i> |
|---------------------|--|---|--|--|
| Putonghua | 太阳 t'ai ⁵¹ • iɑŋ ¹ 老爷儿 lau ²¹⁴⁻²¹ ier ³⁵ 日头 ʒ ⁵¹ • t'ou ¹ | 月亮 ye ⁵¹ • liɑŋ ¹ | 星星 ɕiŋ ⁵⁵ • ɕiŋ ² | 打雷 ta ²¹⁴²¹ lei ³⁵ |
| Hong Kong Cantonese | 日头 jɛt ² t'ɐu ²¹⁻³⁵ 太阳 t'ai ³³ jœŋ ²¹ | 月光 jyt ² kwœŋ ⁵⁵ 月亮 jyt ² lœŋ ²² | 星 siŋ ⁵⁵ | 行雷 haŋ ²¹ lœy ²¹ |
| Shanghainese | 太阳 t'a ³³⁻³⁴ fiā ¹³⁻⁴⁴ 日头 ŋji ^{ʔ1-11} dʁ ¹³⁻⁴⁴ | 月亮 fiyɿ ^{ʔ1-11} liā ¹³⁻²³ 月光 fiyɿ ^{ʔ1-11} kuō ⁵³⁻²³ | 星 ɕiŋ ⁵³ | 雷响 le ¹³⁻²² ɕiā ³⁴⁻⁴⁴ 打雷 tā ³⁴⁻⁴⁴ le ²³ |
| Fuzhou | 日头 ni ^{ʔ544} t'au ⁵³ | 月 ŋuo ^{ʔ5} | 星 siŋ ⁴⁴ | 拍亮线 p'a ^{ʔ23-21} liɑŋ ⁴⁴ s nian ²¹² |

Note: The characters are transcribed in IPA symbols. The superscripted numbers are tone types with 1–5 values.

2312–80 standard met the demands of hardware and software development and exchange of information for general purposes, but it soon had to be amended as new demands arose. In 1994, a standard coding scheme for two supplementary sets consisting of 7237 and 7039 characters was officially announced. As GB 2312–80 was designed to accommodate simplified characters, the new GB 12345–90 was introduced for nonsimplified characters that are maintained in Taiwan and Hong Kong. Nowadays, character recognition for both print fonts and handwriting is commercially available. Text-to-speech synthesis and production in the genre of journalistic texts has achieved a high degree of naturalness. The character script and lexical tones, which were thought to be two major obstacles for Chinese information processing, are no longer condemned, but appreciated as features with a flavor of real Chinese.

Bibliography

- Chao Y R (1968). *A grammar of spoken chinese*. Berkeley, CA: University of California Press.
- Chen P (1999). *Modern Chinese*. Cambridge: Cambridge University Press.
- DeFrancis J (1986). *The Chinese language*. Hawaii: University of Hawaii Press.
- Matthews S & Yip V (1994). *Cantonese: a comprehensive grammar*. London: Routledge.
- Ramsey S R (1987). *The languages of China*. Princeton University Press.
- 傅永和, 1999《中文信息处理》广东教育出版社
- 候精一、梁玉璋、冯爱珍, 1996《福州话音档》上海教育出版社
- 候精一、游汝杰, 1998《上海话音档》上海教育出版社
- 候精一、张双庆、林建平, 1999《香港话音档》上海教育出版社
- 刘涌泉等, 1987《中国语言地图集》Longman

Chinese as an Isolating Language

J L Packard, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

If we use the term ‘isolating’ in what is perhaps its simplest and most often used sense – referring to whether the words of a language are mostly monomorphemic (*see Classification of Languages*) – then Chinese can be considered only a moderately isolating language, because Chinese has at least as many multimorphemic as it has monomorphemic words. The term isolating, however, has also been used to refer to whether the morphemes of a language are clearly identifiable, defined by the following properties: (1) whether morpheme boundaries in the language are sharply defined, (2) whether there is only a single distinct morphemic identity represented within a defined morpheme boundary space (i.e., the extent to which there is no overlapping exponence; *see Classification of Languages*), and (3) whether morphemes in the language have a single, invariant phonological form. If we define an isolating language based on an identifiable morphemes criteria, then Chinese scores relatively high on the ‘isolating language’ scale. It can be profitably studied using both definitions of the term.

Isolating Defined as Having Monomorphemic Words

The definition of isolating language as monomorphemic relies on whether words in a language appear without the obligatory affixation of grammatical morphemic information. This property was intended to contrast with languages such as Russian and Latin in which word roots are generally bound content forms that require affixation of grammatical morphemic information (indicating such properties as case, number, or gender) when they occur in context. For example the Russian root for ‘book’ (*knig-*) must be augmented with an inflectional ending that reflects case or number (*knig-u* book-ACC.SING; *knig-i* book-NOM.PL), and cannot appear as a bare stem in isolation.

Languages like Chinese whose words occur without such obligatory grammatical marking are considered isolating because the words in such languages may appear in bare form without the necessity of adding morphemic information. The absence of obligatory affixation means that words in such languages will tend to contain fewer morphemes on average, giving rise to the monomorphemic word definition of isolating language.

As it turns out, many (if not most) Chinese words are in fact dimorphemic, consisting of either (1) two free content morphemes (compound word), (2) one free and one bound content morpheme or two bound content morphemes (bound root word), (3) a free or bound content morpheme plus a word-forming affix (derived word), or (4) a free content morpheme plus an inflectional affix (grammatical word; *see Packard, 2000* for further details). However, most dimorphemic Chinese words are either compound words or bound root words, and so the multimorphemic status of Chinese words is generally not due to the presence of affixation. Moreover, when Chinese words do contain affixes, they are never obligatory in the sense that they are required in the default case, as seen in the Russian example above.

Chinese affixes are, nonetheless, sometimes obligatory in an alternative sense: if a property in question is selected to be expressed by the speaker, then the use of the affix concomitant with that property is a required element. Some common examples of this obligatory marking of an optionally selected property in Chinese are the use of classifiers with nouns, the marking of plural numbers on human pronouns, and the use of aspect marking on verbs.

Classifiers are word-forming morphemes that are required when nouns are modified by a number and/or a determiner. For example, the noun *shu* ‘book’ generally occurs in context in bare form with no grammatical marking whatsoever. But when *shu* is modified by a number such as *san* ‘three’ or a determiner such as *na* ‘that,’ the classifier *ben* ‘volume’ must occur between the modifying element and the noun, yielding *san-ben shu* and *na-ben shu* for ‘three books’ and ‘that book’ respectively.

In the case of human pronouns, the personal pronouns *wo* ‘I/me,’ *ni* ‘you,’ and *ta* ‘he, she’ are obligatorily marked with the plural suffix *-men* when the referent is plural in number, to yield *women* ‘we, us,’ *nimen* ‘you (pl),’ and *tamen* ‘they, them.’

Verbs in Chinese may occur with inflectional suffixes that express various forms of grammatical aspect, that is, that refer to the activity profile of the event represented by the verb. For example, the verbal aspect marker *-le* (note that this is the *-le* that affixes to and has scope over the verb, and not the *le* that occurs in sentence-final position and has scope over the sentence) indicates that the event associated with the verb has been completed, the verbal aspect marker *-guo* indicates that the event associated with the verb has occurred at least once, and the verbal aspect marker *-zhe* indicates that the action represented by the verb is ongoing or continuous.

In Chinese, the obligatory marking of a selected property as seen in classifiers, human plural pronouns, and verbal aspect contrasts with cases in which the marking of a selected property is optional, as with plural marking on regular human nouns. When a human noun is transparently plural in number, the addition of the suffix *-men*, which would explicitly represent a plural number, is optional. For example, in both of the following examples the Chinese noun that translates into English as ‘teachers’ refers unambiguously to a set that contains multiple members.

- (1) laoshi dou you shu
teacher all have book
‘the teachers all have books’
- (2) laoshimen dou you shu
teacher-PL all have book
‘the teachers all have books’

Both examples refer to ‘teachers’ as a plural concept but only the second overtly marks the plural number with the suffix *-men*. The two examples are identical in meaning, but the second explicitly marks the plural while the first does not.

If Chinese is examined as an isolating language based on its use of monomorphemic words, it is worthwhile to consider in concrete terms where Chinese should be located on the monomorphemic word scale. The contemporary Chinese novel *Shui Ru Da Di* by Wen Fan (2004; Beijing: People’s Literature Publishing House) provides a typical sampling. If we examine the first 100 words in the third paragraph on page 16, we find that 51 (51%) of the words are monomorphemic (if by token; 35 words or 47.2% if by type), 45 (45%) of the words are dimorphemic (if by token; 35 words or 47.2% if by type), and 4 words (4% if by token, 5.4% if by type) contain more than two morphemes. If counted by type, 47.3% of the words are monomorphemic, and 52.7% are multimorphemic.

In addition, the average number of morphemes per word token for that hundred-word sample is 1.54. This figure may be compared with the 1.06 morphemes-per-word cited for Vietnamese (perhaps the most purely isolating language using this criterion), 1.68 for modern English, and 3.72 for Eskimo (see *Classification of Languages*). In sum, if the concept of monomorphemic words is used as the defining criterion, Chinese must be considered only moderately isolating.

Isolating Defined as Having Clearly Identifiable Morphemes

To determine where Chinese belongs on the isolating language scale using the ‘identifiable morpheme’

criterion, the first property to consider is sharply defined morpheme boundaries. In Chinese, morpheme boundaries are nothing if not clearly defined. There is generally no question where one morpheme ends and another one begins in any Chinese utterance. Even in cases of affixation in which the phonological form of the stem is affected, it is quite clear which part of the affixed word belongs to the stem and which part belongs to the affix.

To illustrate, consider the following examples of *-er* (phonetically [ər]) diminution suffixation (data from Cheng, 1973; in IPA, tones not marked). The *-er* suffix often makes only a negligible semantic contribution to the derived word, but it is the affixation operation that has the greatest phonological effect in (Mandarin) Chinese.

The *-er* suffix attaches to words with varying degrees of phonological effect on the stem and on the affix itself. In examples (1)–(3) of Table 1, the *-er* suffix is appended to the stem with the [ə] vowel of the suffix dropped in favor of stem vocalic elements, and with no effect on the phonological form of the stem. In (4), the [ə] vowel of the suffix is dropped and the stem final velar nasal [ŋ] is lost, but its nasality is retained in the form of nasalization on the stem nuclear vowel, that is, [ã]. In (5), the [ə] vowel of the suffix is dropped and the stem final apical nasal [n] is lost, but its nasality is not retained as in (4). In (6), we see a stronger contribution from the suffix, since it retains its [ə] vowel. In (7), the suffix is appended in unaltered form, and the stem final [n] is displaced. In (8)–(10), the suffix is appended in unaltered form, replacing various parts of the stem final, including its complete replacement in (9) and (10).

The examples in Table 1 demonstrate that even though suffixation of *-er* results in a good deal of phonological variability on both stem and affix, in all cases the resulting derived words contain phonological strings that can be unambiguously attributed to either the stem or the affix, and the phonological

Table 1 Some phonological effects of *-er* suffixation

| | Noun | Noun plus <i>-er</i> ([ər]) suffix | Meaning |
|------|------|---------------------------------------|-----------|
| (1) | niou | niour | ‘ox’ |
| (2) | ua | uar | ‘frog’ |
| (3) | kʏ | kʏr | ‘song’ |
| (4) | gaŋ | gār | ‘jar’ |
| (5) | p’an | p’ar | ‘pan’ |
| (6) | i | iər | ‘clothes’ |
| (7) | in | iər | ‘seal’ |
| (8) | kuei | kuər | ‘ghost’ |
| (9) | ci | cər | ‘word’ |
| (10) | pei | pər | ‘cup’ |

Mandarin Chinese has four lexical tones: a high (H) tone, a mid-rising (MH) tone, a low (L) tone, and a high-falling (HL) tone. The L tone sandhi rule changes an L into an MH when the L precedes (i.e., occurs to the left of) another L. The MH tone sandhi rule changes a nonfinal MH into an H when it follows (i.e., occurs to the right of) by an H or an MH. In (3), the citation tones for 'to bury a horse' are MH and L, and their surface realizations are the same as their citation forms. In (4), the tone on the word 'buy' in 'to buy a horse' changes from citation L to sandhi MH

| | | | | |
|-----|-----|------|------|---|
| (5) | fen | shui | ling | <i>divide water mountain-ridge</i> 'watershed' |
| | H | L | L | citation tones |
| | H | MH | L | sandhi tone forms (intermediate, nonrealized forms) |
| | H | H | L | sandhi tone forms (final surface forms) |

To conclude, the reputation of Chinese as an isolating language is perhaps not so well-deserved if we rely merely on the monomorphemic word criterion, since the preponderance of Chinese words are multimorphemic. But if our criterion is how easy the morphemes of a language are to identify and individuate, then Chinese scores rather high on the isolating language scale.

Chen M Y (2000). *Tone sandhi: Patterns across Chinese dialects*. New York/London: Cambridge University Press.

Cheng C (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Packard J L (2000). *The morphology of Chinese*. New York/London: Cambridge University Press.

Choco Languages

D Aguirre Licht, Universidad de los Andes,
Bogotá, Colombia

© 2006 Elsevier Ltd. All rights reserved.

Present Indians of Western Colombia

Colombia, a basically Spanish-speaking nation in the northwestern corner of South America, conserves a considerable number of Indian languages. The speakers of the various languages survived the colonization of the subcontinent, isolating themselves in rather desolate places far from the urban centers, where there were no mestizos and only the black population dared to enter. Of the immense mosaic of aboriginal languages thought to have existed when the Europeans arrived in what is today Colombian territory – for its privileged situation as a crossroads for peoples from north to south and from south to north of the continent – about 90 peoples still survive. They are characterized as different from the majority of the Spanish-speaking population, because they maintain particular sociocultural characteristics, among them a language of their own, as is the case for 65 of these 90 peoples. These languages have been characterized by the most diverse range of linguistic varieties, i.e., isolating, agglutinating, and flexive languages, as correspond to such highly varied regions in which their speakers are found: desert zones, grasslands, jungles, coastal littorals, river littorals, foothills, and mountainous zones of both temperate and cold climates.

Four of these Indian languages still survive in western Colombia, which correspond to four ethnic groups that continue to preserve their own cultural characteristics, such as their language and, therefore, their particular way of thinking, or worldview. The first of these four languages is Tule, of the Chibchan linguistic family, the speakers of which are known as 'Cunas.' They occupy the extreme northwestern part of the country, in the Golfo de Uraba (where there are no more than 1000 individuals) and the majority are found in the neighboring country of Panama, in the San Blas Islands (around 40 000 individuals), where they have immigrated for more than half a century. The second language is Awa or Awa-Cuaiquer, classified as an independent language, whose speakers are thought to number about 4000 and are located in the extreme southwestern part of the country (in the department of Nariño) and in smaller numbers in the neighboring country of Ecuador. The third and fourth languages are Waunméu (Woun Meu) and Embera, which belong to the so-called Choco language group, which has only recently been classified as an independent linguistic family.

Waunméu is spoken by the Waunanas, who number around 4000 individuals in Colombia, along the lower San Juan River, in the south of the department of Choco, and no more than 2000 individuals who have immigrated to the Province of Darien, in Panama. Embera is spoken by the Indians who call themselves Emberas but who are known by different names in the literature because they constitute a much larger number of speakers – around 60 000 – divided in various dialects. The Emberas are dispersed throughout the western part of Colombia and even in the frontier zones of Panama and Ecuador, and some of these dialects have grown so far apart that they are now mutually unintelligible.

Of course, this is a timid sample of the much greater number of ethnic groups that inhabited the western part of Colombia when the Europeans arrived, among which we can recall the names of the Idabáez, Ingarás, Birus, Surrucos, Poromeas, and the present-day Kunas, Waunanas, Katíos, or Emberas. Some of the denominations applied to the Indians then generally known as 'Choco' or 'Chocoes' were the Andáguedas, Baudó, Chamís, Dabeibas, Dariens, Katíos, Noanamás, and Saijas. Nowadays it is known that these names are derived from the names of the regions inhabited by these groups, which generally took the name of the main river that crossed through their territory and, in the case of the name 'Katío,' to the fact that the Embera Indians eventually occupied the region of the Katío Indians, a brave warrior tribe that succumbed to the Spanish.

The Embera Indians occupy a much greater territory today than they did at the time of the arrival of the Europeans, but with a very atomized coverage, i.e., only in different and specific points of little extension. Mestizo settlers displaced them to these Indian reservations, called 'Resguardos' or 'Cabildos,' which were very effective in the colonial period in preventing the extinction of these peoples, by impeding their occupation by outsiders but obliging the Emberas to give up the extensive territories in which they had once freely roamed.

In this article we see the different dialects into which the Embera language is presently divided. These dialects are a product of the different regions in which the Embera Indians have settled since the arrival of the Spanish, in different latitudes of the continent but always limited to a fringe that extends from the western littoral: the Pacific coast of Colombia, from north to south, to the Cauca River, which separates the western and central cordilleras stretching from north to south along the country, together with the eastern cordillera, the final branches of which

disappear as they enter the Caribbean region of Colombia. Thus, the scenario that the Choco Indians occupy consists of the Pacific coast of Colombia, with its jungle plains; the Province of Darién, in Panama; and the spurs of the western cordillera and its terminal branches to the west of the Cauca River.

Retrospective of Linguistic Studies and Attempts to Classify the Emberas

The inclusion within a single linguistic family of the speech of the different Choco groups (the Waunana language and the different Embera dialects) that survived the Conquest and the colonial period is a recent fact. Their classification within any one of the great variety of American linguistic families is still open to discussion.

In the literature on the country's Indians, there is abundant documentation on population and migrations of the Choco, from chroniclers like Fray Pedro Simón, Bartolomé de Las Casas, Jorge Robledo, Juan de Castellanos, Pedro Cieza de León, to recent researchers like Henry Wassen, Katleen Romoli, Reina Torres de Arauz, Sven Isacson, Mauricio Pardo, and Patricia Vargas. The last two, who are Colombian authors, have advanced in research about the Emberas, having reviewed all previous studies. In his article 'Bibliografía sobre indígenas Choco' (1981), for example, Pardo did an excellent review of the ethnohistoric literature available to date, and in 'Regionalización de indígenas Choco' (1987), he updated the discussion of the ethnohistoric panorama. Vargas (1986), on the other hand, found that the incursion of the Emberas into the territories of the Katio Indians did not mean the total extinction of the latter, because the two peoples intermingled, which is why the present Emberas of the region present particular characteristics that could be assigned to the Katíos.

The term Choco was already used in the 17th century to designate the Emberas of the upper San Juan and Atrato rivers and the Waunanas of the lower San Juan River. The earliest report known about the Emberas is found in the diary of the missionary Father Joseph Palacios de la Vega, around 1787, in San Cipriano, on the San Jorge River. This linguistic material, consisting of 37 phrases and 107 morphemes, fundamentally corresponds to the speech of the present Emberas of the northeast (Reichel-Dolmatoff, 1955).

A series of vocabularies was later collected by travelers, mostly foreigners, in different Choco Indian localities (Mollien, 1824; Cullen, 1851; Seeman, 1851; Bastian, 1876; Greiffenstein, 1878; Collins, 1879; White, 1884; Peláez, 1885; Etienne, 1887; Simons, 1887; Pinart, 1887; Velásquez, 1916; Robledo, 1922). These materials fundamentally

served as the basis for analysis and classification until the middle of the 20th century.

But there have also been comparative studies since the 19th century: Bollaert (1860) proposed affinities between the Choco and Mesoamerican groups; Adam (1888) compared vocabularies obtained by Cullen, Seeman, and Uribe; Brinton (1891) observed the territorial extension of the speech of the Choco; Chamberlain (1907) determined that the geographical limits of the Choco were between 8 and 4 degrees northern latitude, between the Golfo de Urabá and the Golfo de San Miguel, and proposes Choco as an independent linguistic group; Lehmann (1910, 1920) suggested kinship with the Chibcha dialects of the Barbacoas and Talamasca groups; Loukotka (1968 [1942]) reaffirmed the separation of these languages as an independent linguistic family and recognized nine extant languages and five extinct languages; Rivet (1912, 1924, 1943) compared elements of the Choco vocabulary with 56 Caribe dialects, 34 Chibcha dialects, and 29 Arawak dialects and concluded that there was a strong Caribe influence and, to a much lesser degree, Chibcha and Arawak influence; Ortiz (1937, 1940, 1954, 1965), Mason (1950), Meillet and Cohen (1952), and Tovar (1961) followed Loukotka's regionalization and Rivet's affiliation.

The first attempts to classify the native languages of America were made in the second half of the 20th century. At the beginning of the 20th century, 19 independent language families were mentioned for the Pacific coast, including the Choco family (see, for example, the classifications of Alexander Chamberlain [1913] for the linguistic families of South America). Later researchers, such as Paul Rivet (1944), reduced this number and proposed the inclusion of the Choco family within other macro-families, like the Chibcha or the Caribe. At present, in light of recent linguistic explorations, the thesis of the independence of this family seems to be the most reliable, vindicating its defenders, among whom, in addition to Chamberlain, we can name Nordenskiöld (1928), Loukotka (1968 [1942]), Tovar and Larrucea (1984), and Pardo and Aguirre (1993).

The cultural unity and the common origin of the Choco Indians were the subject of controversy for a long time. Mason's classification (1950) (broadened with that of Greenberg, 1960), for example, divided the Choco languages into Empera, with 3 variants; Catío (Embera-Catío), with 14 variants; and Noanamá (with 1 variant). Loukotka (1968 [1942]) had already spoken of 9 extant and 5 extinct Choco languages, and later Loukotka and Rivet proposed 10 extant variants for the Choco group (which they call the Emperá division) and 2 extinct ones (see Ortiz, 1965: 197–200).

Jacob Loewen (1960) confirmed Erland Nordenskiöld's statement, testifying that linguistically only two languages (Waunana and Embera) – mutually unintelligible but nonetheless related – belonged within the Choco family, and proposed, with a phonological criterion, four large dialectal areas, one Waunana and three Embera, with lexical variations within the Embera areas.

Among the main bibliographical compilations on the Choco languages were those of Adam (1888), with 7 references; Lehmann (1920), with 30 references; Reichel-Dolmatoff (1945), with 38 references; Ortiz (1954), with 60 references; Loewen (1963), with 191 references, which were not only linguistic but historical as well; Ortega (1978), with 67 references; and Pardo (1981), with 72 references, and (1986), a survey of everything written on the subject to date, with 135 references, including everything from academic studies to simple lists of words.

There have been grammatical studies of the Embera language since 1881, when José Vicente Uribe published a brief article in which he presented the different types of Embera words in a general way. In 1936, Fray Pablo del Santísimo Sacramento published a grammatical essay on the speech of the Embera-Catíos of the Apostolic Prefecture of Urabá, as well as a classification of Embera, in which he dedicated a small portion to the syntax of the language. In 1918, an anonymous Catío-Spanish catechism appeared for missionaries of Antioquia. There is also an undated Catía grammar by María Betania (quoted in Pinto, 1974), and the Claretian priest Constancio Pinto published a Catío-Español dictionary (1950), as well as another extensive dictionary with grammar (1974).

Scientific studies based on fieldwork began with the research of Jacob Loewen, an American Mennonite missionary who did a study for a master's degree (1954) among the Waunana Indians of the lower San Juan River, and a doctoral thesis (1958) on the speech of the Emberas of the Sambú River, in the province of Darién in Panama. Loewen also wrote numerous articles on Embera phonology and dialectology, on comments on traditional stories, on loans from Spanish, on problems of bilingual literacy programs, and on basic readers in Indian languages.

Jean Caudmont (1955) elaborated notes on phonological and grammatical generalities through the use of field notes of Reichel-Dolmatoff (1945), taken 10 years earlier among the Embera group in Riofrío in the department of Valle, who had emigrated from the region of the Chamí. The Claretian missionary Constancio Pinto, who lived with the Emberas of the region of the Chamí (headwaters of the San Juan River) for more than 40 years, published a dictionary (1950) of the Embera language, as well as a book with a much more extensive vocabulary and with a

section on grammar (1974). Despite their having been based on methodical fieldwork, these studies suffered from the fact that they had been transcribed using Spanish-language phonetics and indistinctly presented, especially in the work of Pinto, words from zones like the Chamí, the Andágueda, the Sinú, and the Atrato, without taking the dialectal variations into account. With more linguistic precision, the Swedish researcher Nils Holmer (1963), in one of the publications of the Gutemberg Ethnographic Museum, occupied himself extensively with phonological and morphological aspects of Waunana.

The Summer Institute of Linguistics (SIL), which arrived in the country in 1962, carried out linguistic studies in distinct zones inhabited by Embera Indians. Francés Gralow (1976) elaborated a phonological description for the Chamí zone. In the 1970s, Eileen Rex and Mareike Schotlenndreyer traveled throughout the municipalities of Dabeiba, Frontino, and Chigorodó, in the department of Antioquia, and the upper Sinú, in the department of Córdoba, and published a phonology (1973) of the speech of the Emberas of the upper Sinú and northwestern Antioquia. Schotlenndreyer developed a literacy primer for the zone of Chigorodó (1973) and a structural analysis of her and Rex's stories (1977). Eileen Rex wrote her master's thesis on the Catía grammar (1975). Phillip Harms developed basic readers on the Embera language and tales and stories in the company of the natives from 1981 to 1985, for the Emberas of the Saija River on the coast of the department of Cauca, to the south of the department of Choco, and carried out a phonological description with Judy Powell (1984) and a grammatical study of the speech of these Emberas (Harms, 1987). Powell also mimeographed some Embera stories and biblical passages. David Stansell, who lived for more than 10 years among the Emberas of the Bojayá River in the department of Choco, wrote about them in *Aspectos de la cultura material de grupos étnicos de Colombia* (1973). Michael and Nellis (1984) produced primers in Chamí for the Emberas of the Valle de Garrapatas in the department of Valle del Cauca.

Gordon Horton worked with the Emberas of the upper Sinú in the 1960s and 1970s, mainly on the morphology of the language, and developed a series of primers and other didactic materials. Miguel Loboguerrero carried out a linguistic study (1976) on the dialect of the Chamí region; his resulting work included a phonological description, a grammatical description, and a corresponding lexicon. Nelly Mercedes Prado did an analysis of the 'Epera' variant ('Embera,' according to the phonology of this dialect) of the Saija River (1982) as her master's thesis, the presentation of which included phonology, morphology, and

an appendix titled 'Un estudio inicial,' along with a lexicon of 845 items, each with its respective phonetic transcription. She continued her work with the publication of didactic materials (1985), further explored aspects of the language, such as nasality (1991), and later worked in ethnolinguistic conflicts between blacks and Indians (1992) within the broadest project, known as 'Cada río tiene su decir.'

Many missionaries working in Embera territories have concerned themselves with the language. One primer on the language of the Emberas of the upper San Juan, with an alphabet, was developed by G. Manzini (1973); another primer, on the Katía variety, was designed by Martínez and Guisao (1980). There are a catechism in the Baudó dialect (1981) and a primer by Livia Correa (1982), as well as one, by María L. Picón (1985), on the Itsmina region.

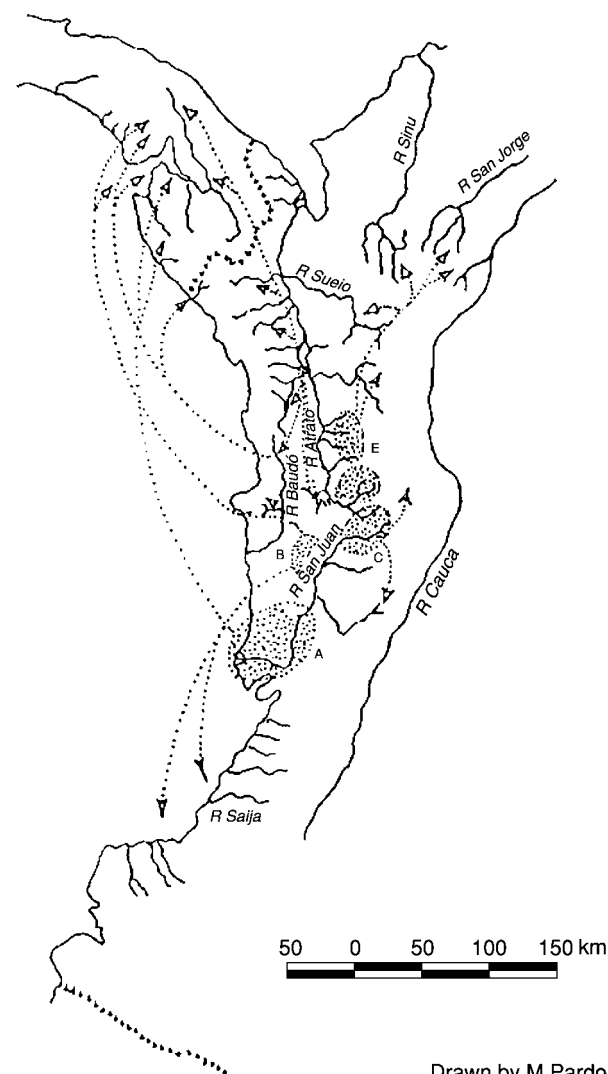
For the Waunanas, in addition to Holmer's studies, there was a phonological and grammatical study done by the Sacred Heart missionaries Sánchez and Castro (1977), with the advice of Reinaldo Binder of the SIL, and a monograph by Luz Lotero (1972).

The Embera Waunana Regional Organization (OREWA) of Choco wrote a manual for indigenous teachers (1987), within the framework of its newly initiated ethnoeducation program.

Mauricio Pardo has done phonological and grammatical descriptions of the Embera language in north-western Antioquia and the zone of the upper Baudó River in the department of Choco. With his participation in workshops with teachers from Baudó and in 1983 in northeastern Antioquia, an era of studies began that committed both Indians and researchers to a common cause in the application of the results of linguistic studies. In 1986, Pardo proposed, together with the author of this article, a revision of the Choco dialectology established by Loewen, 1963 (see next section of this article), and has done an extensive compilation of the publication of linguistic data on this language up to 1986. This author has also concerned himself with the elaboration of language primers and sociolinguistic aspects of the ethnic group.

Present Regionalization of the Embera Indians

The first Indians denominated 'Chocos' by the Spanish were the Emberas of the upper San Juan River, who were then known as the Simas or the Tatamás. These Indians today call themselves Chamí. This denomination would later be applied to all indigenous groups of the upper Atrato River, in the department of Choco, then known as 'Citará' or 'Citarambirá,' and to the Indians of the middle and lower San Juan, respectively called 'Poya' and 'Noanama' in



Choco denominations in the time of the Conquest

- A: Noanama (Lower San Juan)
- B: Cirambirá, Poya (Mid San Juan)
- C: Tatama, Sima (Upper San Juan)
- D: Citara (Upper Atrato)
- E: Eastern tributaries of the Atrato River

----> : Post-Colombian Migrations

Figure 1 Current Choco Dialectology. Reproduced from Pardo M (1987). 'Regionalización de indígenas choco.' In *Revista del Museo del Oro*, Boletín 18, January–April. Bogota: Musco del Oro. 46–63, with permission.

the the 17th century. Based on these points, registered in colonial papers, and respecting the linguistic data obtained from present settlements, one can attempt to reconstruct the dispersion of the Chocos (see Figures 1 and 2).

Most of the Chamí are located along the upper San Juan River, in the Risaralda municipalities of

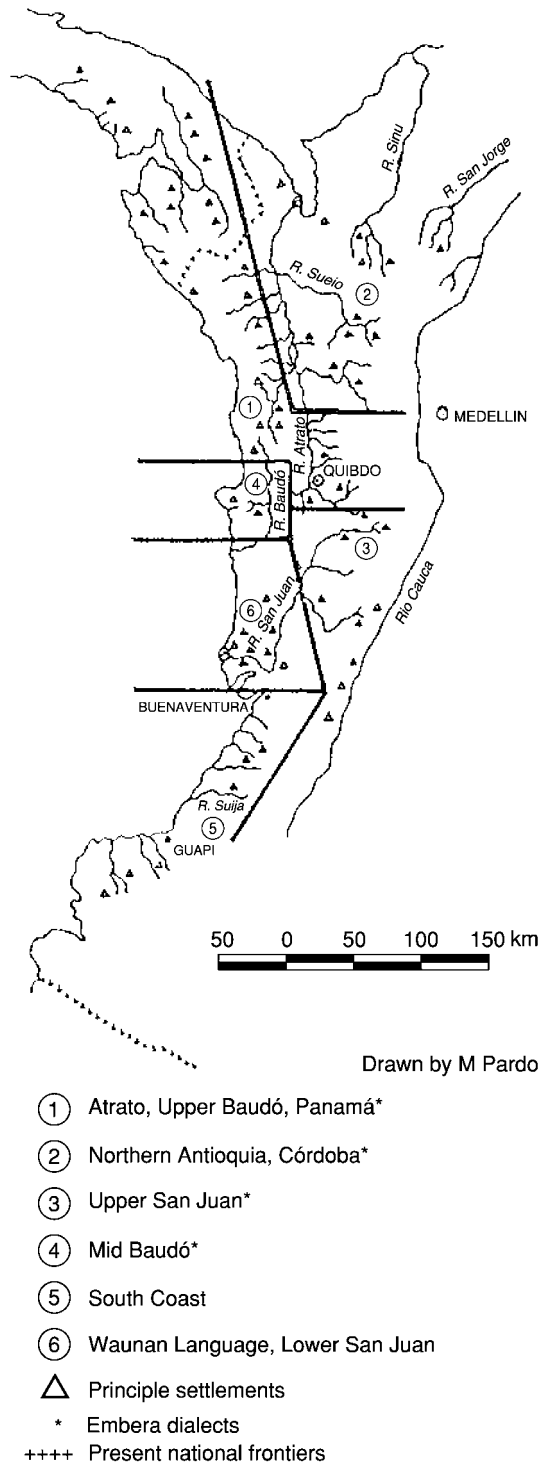


Figure 2 Choco Dispersion. Reproduced from Pardo M (1987). 'Regionalización de indígenas choco.' In *Revista del Museo del Oro*, Boletín 18, January–April. Bogotá: Musco del Oro. 46–63, with permission.

Mistrató and Pueblorico, on the border with Choco. They have moved northward and southward along the cordillera to places like the upper Andágueda River, in southeastern Choco, to the southwestern

part of the department of Antioquia in the municipalities of Jardín, Valparaíso, and Bolívar, and to the northern part of the department of the Valle del Cauca along the Garrapatas and Sanguinini rivers. Small groups are also located in other parts of Antioquia and Valle and have even moved down into the departments of Caquetá and Putumayo.

Those who were called Citarás or Citarambirás during colonial times – then located along the upper Atrato River, on the Capá River, in Lloró, and along the lower Andágueda River – have moved northward along the river to the upper Baudó River, toward the coastal tributaries to the north of Cabo Corrientes and the Panamanian portion of Darién. These river-dwelling Indians are known as 'Cholos' on the Pacific coast of Colombia.

Because these people form a distinct dialectal zone and because they are generally considered a mountain group, researchers believe the Indians who presently occupy territories in northeastern Antioquia – in Dabeiba, Frontino, Ituango, Murri, among other places, and in the department of Córdoba, in the upper Sinú, San Jorge River, Rioverde, etc. – must descend from Emberas who, after the Conquest, settled along the eastern tributaries of the middle course of the Atrato River, a group different from the Citarás. These Indians are erroneously known as 'Katíos,' but colonial documents imply that the real Katíos succumbed toward the end of the 17th century, after a terrible struggle with the Spanish. Vargas (1990) postulated, based on archival documents, that many Katíos united both in alliance and in war with the Emberas.

The Indians encountered by the Spanish in the middle San Juan River, whom the Spanish called 'Poyá,' are believed to be the ancestors of the present dwellers of middle Baudó River, in the affluents Catrú, Dubasa, and surroundings. The Poyá presented a dialectal difference with the ones from the upper Baudó River. These people called themselves Emberas to differentiate themselves from the mountain people, who were called Katíos.

The Indians presently located to the south of Buenaventura also descended from the Poyás, whose main settlements are along the Saija River (department of Cauca), and the Satinga and Saquianga rivers (department of Nariño) (Pardo, 1987). They call themselves 'Eperas,' in accordance with the phonology of their dialect.

In the department of Caldas, there are settlements of Embera Indians, known to the rest of the population as 'Memes.' They live in municipalities such as Belalcázar, Vitervo, and Riosucio, in places like La Betulia, La Tesalia, and the Indian reservations of San Lorenzo and Nuestra Señora de la Montaña. Some are Indian reservations with reserved territory, while

others such as Cañamomo and Lomapieta are in the process of becoming reservations (these are called ‘partialities’). In addition to the problem of vindicating their own identity as a separate ethnic group, they have encountered major difficulties for having lost their native tongue, but nonetheless they are at present actively committed to carrying out programs to recover their language with the help of native speakers from other regions.

The Emberas who settled along the lower San Juan River and its tributaries, along the Juradó, Jampavadó, Docampadó, and Siguirisúa in southern Choco, and along the San Juan de Micay River in Cauca were called ‘Nonamá’ or ‘Noanamá’ ever since the invasion, but they call themselves ‘Waunana’ or ‘Waunán.’ Over the course of a century they have migrated to the province of Darién in Panama, where 2000 now reside, and to the Chintadó River along the lower Atrato, where there are several hundred who migrated some 20 years ago. There are estimated to be about 4000 native speakers of Waunana in Colombia. Like the Emberas, they are known as ‘Cholos.’ The Waunanas and the Emberas are the only two ethnic groups that can clearly be identified as presently forming part of the Choco family.

In 1988, the author of this article, together with the anthropologist/researcher Mauricio Pardo, presented a proposal for the regional classification of the Choco Indians – a revision of that proposed by J. Loewen – based on the different dialects encountered during fieldwork in the different zones with Choco Indians

in Colombia. Some samples to support this proposal are presented below. These are taken from personal fieldwork notes and first appeared in an article entitled ‘Dialectología Choco’ in the memoirs of the seminar-workshop ‘Estado actual de la clasificación de la lenguas indígenas de Colombia,’ held in February 1988 at the Instituto Caro y Cuervo in Bogota (Pardo and Aguirre, 1993) (see Figure 3).

To begin, a diagram showing the present linguistic variations and the local denominations is presented (Figure 3). The proposal of Jacob Loewen is then presented (Table 1), followed by the Pardo-Aguirre proposal (Table 2). After that, the zones and specific places identified by Pardo and Aguirre are presented in detail (Table 3), along with a global diagram of said zones (Figure 4). Finally, phonological and grammatical comparisons of the Waunana language and the different dialects proposed for the Embera language (as well as among the latter) are shown (Tables 4–6).

Present State of Studies on the Embera Language

Colombia, together with the other Latin-American countries, with all the richness that multiculturalism and plurilinguism represent, only in recent times has given attention to its aboriginal languages. There is not still an official position on the defense of these languages and their speakers, who are not extinct, thanks to their proper fight and the support of a sector of the civil population. Just during the

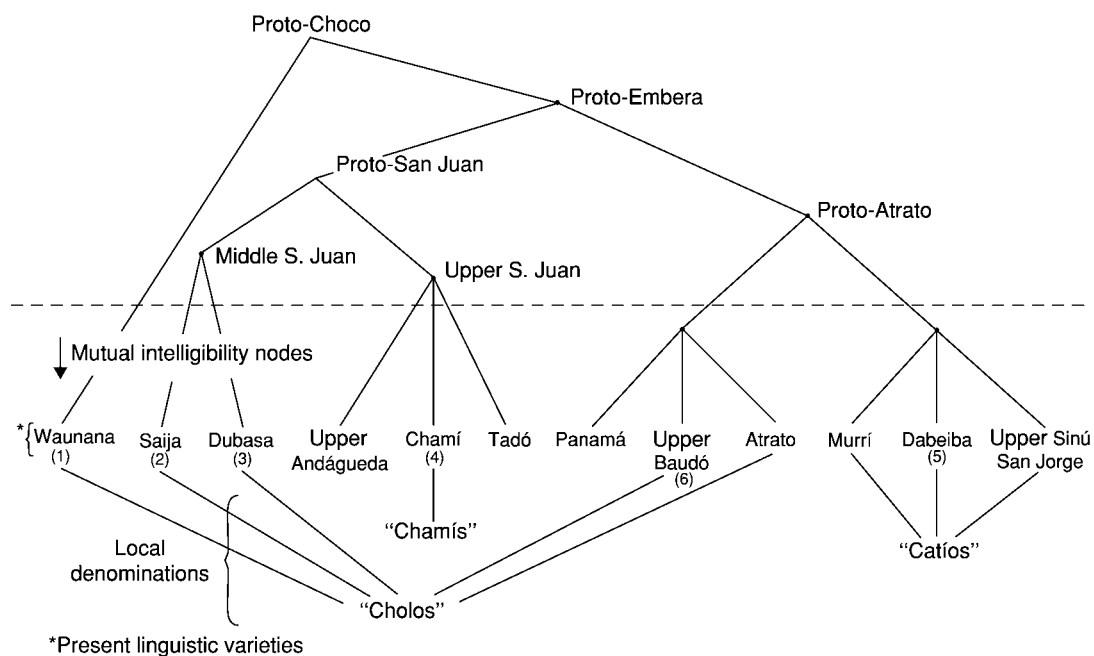


Figure 3 Phylogenetic tree of the Choco linguistic varieties (local denominations).

Table 1 Choco phonological systems (according to Jacob Loewen)

| | <i>Waunana</i> | <i>Saija Baudó</i> | <i>Riosucio Tadó Chamí</i> | <i>Catío-S. Jorge Río Verde Sambú</i> |
|---------------------------|--|--|------------------------------------|---|
| PLOSIVE | | | | |
| Voiceless aspirated | p ^h t ^h k ^h | p ^h t ^h k ^h | | |
| Voiceless non-aspirated | p t k ? | p t k | p t k | p t k |
| Voiceless tense | | | p' t' k' | p' t' k' |
| Voiced | b d g | b d g | b d g | b d g |
| FRICATIVE | | | | |
| Voiceless strong | s š | s č | s č | s č |
| Voiced mild | | | | z ĵ |
| LATERAL | l | l | l | l |
| TRILL | ř r̄ | r ř | r ř | r ř |
| Voiced | r r | r r | r r | r r |
| NASAL | | | | |
| Voiced | m n | m n | m n | m n |
| APPROXIMANT | w y h | w y h | w y h | w y h |
| VOWELS (for all dialects) | | | | |
| Oral and nasal | | | i ī u e o a | |

Note: As can be seen, Loewen proposed 4 phonological systems and dialect subdivisions at the lexical level within them. Nonetheless, the recent data show that at least 6 different systems can be identified: 1 for Waunana and 5 for the different Embera dialects.

Table 2 Choco phonological systems (based on recent data)

| | <i>Waunana^a</i> | <i>South Coast^b</i> | <i>Lower Baudó^c</i> | <i>Upper San Juan^d</i> | <i>Antioquia Córdoba^e</i> | <i>Upper Baudó Atrato Panama^e</i> |
|------------------------------|--|--|--|--|--|--|
| PLOSIVE | | | | | | |
| Voiceless strongly aspirated | p ^h t ^h k ^h | p ^h t ^h k ^h | p ^h t ^h k ^h | | | |
| Voiceless mildly aspirated | | | | p ^h t ^h k ^h | p ^h t ^h k ^h | p ^h t ^h k ^h |
| Voiceless non-aspirated | p t k ? | p t k ? | p t k | | | |
| Voiced tense | | | | b d | b d g | b d g |
| Voiced relaxed | b d g | b d g | b d | | | |
| IMPLOSIVE | | | ɓ ɗ | ɓ ɗ | ɓ | ɓ ɗ |
| AFFRICATE | č | č | č | č | č ĵ | č ĵ |
| FRICATIVE | s h | s h | v s h | v s h | v s z ð h | v s z h |
| LATERAL | l | l | l | l | l | l |
| TRILL | r rr | r rr | r rr | r rr | r rr | r rr |
| NASAL | m n | | m n | m n | m n | m n |
| APPROXIMANT | w j | w j | w j | w j | w j | w j |
| VOWELS (for all dialects) | | | | | | |
| Oral and nasal | | | | a e i o u ʉ | | |

For the South Coast, there is a sixth vowel, which is the ə (oral only)

^aData from Mejia (2000b)

^bData from Prado (1991)

^cData from Pardo (1985a)

^dData from Aguirre (1995a)

last 20 years of the 20th century were the Indian and Afro-Colombian languages, still alive in the national panorama, taken seriously by academia.

In 1984, the Anthropology Department of Andes University instituted a Masters in ethnolinguistics, with the sponsorship of the Centre Nationale de la Recherche Scientifique (CNRS) of France. In the program, researchers are prepared for the study of

the native and Afro-Colombian languages, their eventual goals being publication, conservation, and strengthening of these languages. The program's students and professors constitute the Centro Colombiano de Estudios de Lenguas Aborígenes (CCELA), through which they do the scientific work of the rescue and fortification of these languages. With these linguist students, a new era in the research and

Table 3 Details regarding the zones and specific places of the Choco proposed by Pardo and Aguirre

| | <i>Waunana</i> | <i>Lower Baudó</i> | <i>Upper San Juan</i> | <i>Choco</i> <i>West Atrato</i> <i>Antioquia</i> <i>East Atrato</i> |
|-------------|--|--|---|--|
| PLOSIVE | p ^h t ^h k ^h p t k b d g | p ^h t ^h k ^h p t k b d ɓ d' | p ^h t ^h k ^h b d ɓ d' | p ^h t ^h k ^h b d g ɓ d/ð z j |
| AFFRICATE | | | | |
| FRICATIVE | | | v | |
| TRILL | | ɕ s h | | |
| SOUNDING | | r rr | | |
| APPROXIMANT | | l m n w j | | |

Note: According to this scheme, at the strictly phonological level, Saija and Waunana have identical systems, even though they are very different at the lexical level.

promotion of the aboriginal and creole languages of the country has begun, with them covering the entire national territory, doing fieldwork and linguistic data analysis *in situ*. This has yielded an awakening of these communities for the rest of Colombian population and even for themselves.

Several students from the program have done research on the Embera language:

Rito Llerena Villalobos. He was a student from the first promotion, having finished the program in 1987. He is now a professor at Universidad de Antioquia, in the Department of Linguistics. From 1989 to 1992, he worked on compared phonology of the Amerindian languages of Antioquia, including the Tule language (of the Cuna Indians), subject of his degree thesis (1987). This researcher has worked lately on the Embera language, creating didactic materials for the Indian teachers of Alto Andágueda, phonological and morphological research in the Embera Reservation of Jaidukamá, department of Antioquia, and collaborating in ethnoeducation among the Emberas of Tierralta, upper Sinú River, in the department of Córdoba, where he is working at present. In the year 2000 he wrote a report on the grammar and phonology of the Tule language for the Instituto Caro y Cuervo.

Mario Hoyos Benites. He, too, was a student from the program's first promotion and finished in 1987. At present he is a professor at the Universidad de la Guajira. He worked in 1984 in the Napipí and middle Atrato rivers and other places in the region. His research has addressed everything from the design of didactic material for Indian teachers around all the country (1991) to interdialectal phonology. He presented a report on the Embera language for the Atlas Etnolingüístico de Colombia of the Instituto Caro y

Cuervo in 1997, and wrote a report (2000) on the Embera language of the Napipí River for the institute.

Ernesto Llerena García. He completed the program in 2001, with a dissertation titled 'La predicación de la oración simple en la lengua embera del Alto Sinú' (simple sentence predication of the Embera of Atto Sinú). He has been profesor of linguistics at the Antioquia and Córdoba universities, where is working at the moment. With his father, Rito Llerena, and the Emberas of upper Sinú River, he wrote *Diccionario etnolingüístico de la lengua Embera* (2003) for the Normal Superior de Montería (capital of the department of Córdoba).

Daniel Aguirre Licht. A student from the second promotion, he finished up in 1989. In 1985, he began phonological studies of Chamí, southeast of the department of Antioquia. He continued with morphological studies in 1987, and then morphophonological and grammarians in 1998. In 1988, he collaborated with the anthropologist Mauricio Pardo in research on Choco dialectology; included in the resulting article (Pardo and Aguirre, 1993) was an answer to Paul Rivet's hypothesis about the origin Karib of the Choco languages. Aguirre Licht also worked in the department of Risaralda, the location of the Indian reservation Embera-Chamí of Purembará (from *puru* = 'town' and *embera*), a possible place of the dispersion of the different Embera groups at the Spanish arrival, and he has also worked with the Emberas of Garrapatas Valley, in the department of Valle del Cauca.

About the Waunana there are also the works of Gustavo Mejía (1987), another student from the first promotion of the master's in ethnolinguistics of the Universidad de los Andes. He did a grammarian investigation in 1987 as his thesis. He also did, for the

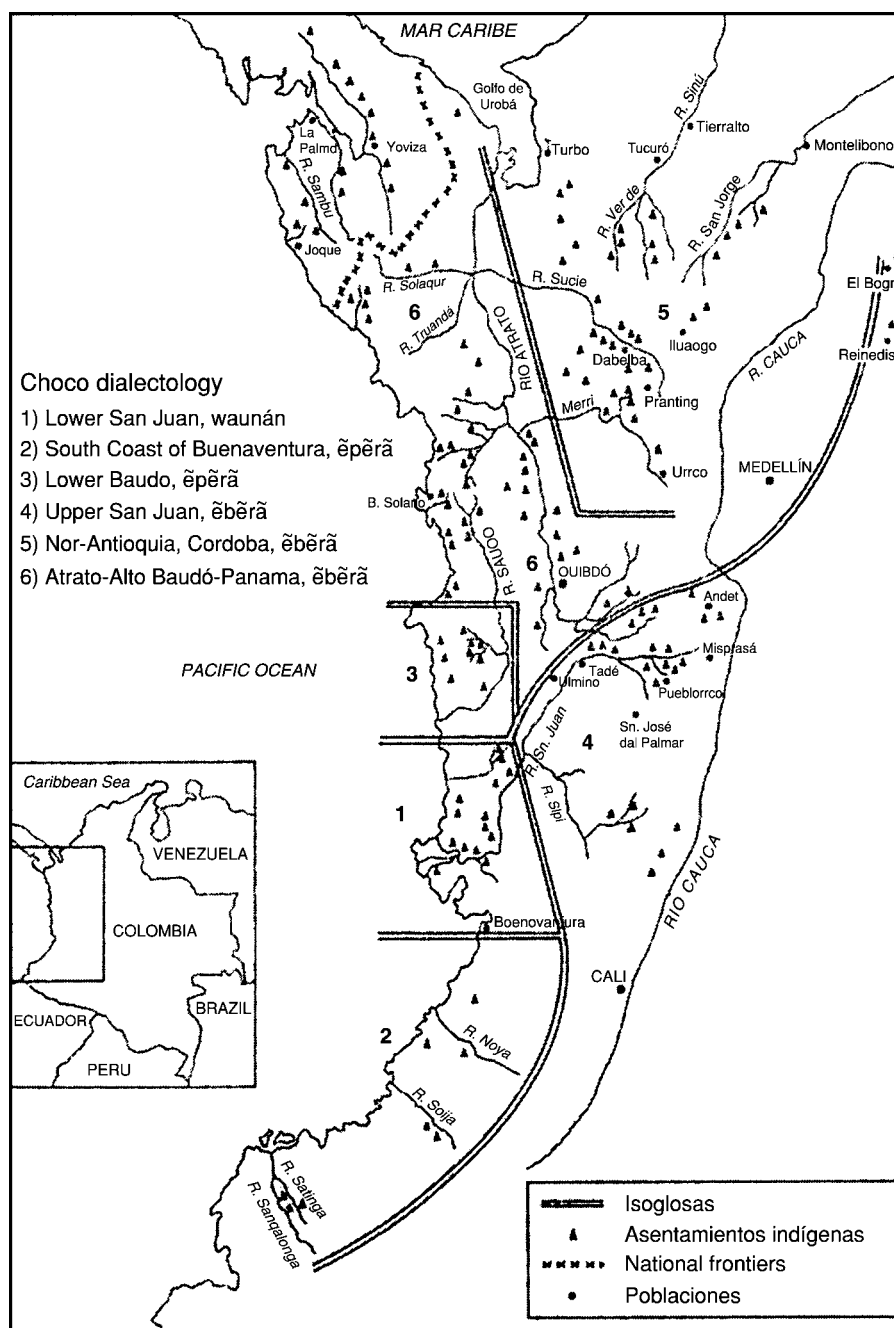


Figure 4 Choco dialectology.

Instituto Caro y Cuervo, a phonological and morpho-syntactic description of Waunana (2000b) and a presentation of the aboriginal languages of the Pacific coast of Colombia (2000a).

Edel Rasmussen, who worked at the Universidad Nacional de Panamá, studied the Embera language of the Panama area and published research on phonology (1986) and on grammar (1985).

The Technological University of Pereira (UTP), located in the capital of the department of Risaralda, has paid attention to the great number of the Embera-Chamí Indians who live in the department, both in

studies of their language and in projects on other matters. Fernando Romero L. of the Psycho-Pedagogy Department in the School of Education does research on linguistic and pedagogical problems of the teaching of Spanish as a second language with bilingual Chamí and Nasa (Páez) teachers, as well as on discourse analysis of this variety of the Embera language, including studies in which the author of this article has participated. Linguist Olga L. Bedoya works with him, and as Director of the Ethnoeducation and Community Development Program in the same school she does research on the interference of

Table 4 Phonological variation according to Lexicon. Representative sample

| | <i>Waunana</i> | <i>South Coast</i> | <i>Lower Baudó</i> | <i>Upper San Juan</i> | <i>Antioquia Córdoba</i> | <i>Atrato</i> |
|----------|----------------------|----------------------|-----------------------|-----------------------|------------------------------|---------------|
| I | mu | mũ | mũ | mũ | mũ | mũ |
| You | pu ^h | pu | pu | bu | bu | bu |
| He | ič | iči | iči | iči | iji | iji |
| We | mač | tai | tači | dači | dai | dai |
| You | paan | pará | mārā | mači/mārā | mārā | mārā/pārā |
| They | hakún | āči | āči | āči | āji | āji |
| Who | k ^h ai | k ^h ai | k ^h ai | kai | kai | kai |
| Person | waunán | ēpērā | ēpērā | ēbērā | ēbērā | ēbērā |
| Man | emk ^h oi | ūmūkūrā | ūmūk ^h irā | mūkīrā | mākīrā | mākīrā |
| Woman | ui | ʔawera | ūērā/vērā | ūērā | ūērā | ūērā |
| Father | ai | ākōrē | tata | čača/dadá | zeze | zeze |
| Mother | at/tata | nāvē | nana | dana/nāvē | papa | papa |
| Son | ieuá | oarra | uarra | varr/oarra | vuarra | oarra |
| Daughter | k ^h a | k ^h au | k ^h au | kau | kau | kau |
| Spouse | huuja | k ^h ima | k ^h imá | kima | kima | kima |
| Head | puru | poro | boro | boro | buru | boró |
| Eye | dau | tau | tau | dau | dabú | dau |
| Tooth | k ^h ier | k ^h ida | k ^h idā | kidā | kiđá/čiđá | kiđá |
| Mouth | i/ihure | it ^h ai | it ^h ae | i/itae | itae | itae |
| Stomach | bi | bi | fi | fi | fi | fi |
| Hand | húa | húa | húa | húa | huwá | huwá |
| Foot | bui | buru/hīrú | hīrú | hūrú/hērú | hīrú | hērú |
| Blood | bak | wáa/iwá | va | oa | va | oá |
| Meat | nemekmót | čier | čik ^h o | kiuru | jiko | jiko |
| Water | du | panía | panía/paitó | banía | banía | baidó |
| Ground | hēp | joró | joró | éoro | egoró | egoró |
| Stone | mok | māū | mōkará | mokara | mōgará | mōgará |
| River | du | to | to | do | do | do |
| Mountain | duursi | ʔee | eja | ea | katumá | ejá |
| Sun | edau | ākōrēhīrú | umāda | umāda | īmāda | umāda |
| Tree | pabú | pak ^h uru | pakurú | bakuru | bakuru | bakuru |
| Leaf | k ^h iri | k ^h iru | k ^h itúa | kidúa | kitúa | keduá |
| Root | pa ^h are | k ^h arrá | k ^h arra | karr | karrá | karrá |
| Dog | saak | usa | usa | usa | usá | usá |
| Bird | nemčai | ipana | īpaná | ibana | ībaná | ībaná |
| Fish | āwarr | cik ^h o | betá | beda | bedá | bedá |
| One | aʔpai | aba | aḃa | aḃa | aḃa | aḃa |
| Two | daunumí | ome | ōmē | ome | ume | umé |
| Three | t ^h arhūp | ōpé | ōpea | ōbea | ūbea | ubea |

Notes: Details regarding the zones and specific places of the Choco proposed by Pardo and Aguirre. **Waunana:** lower San Juan River, Docampadó, coastal rivers, Juradó, Panama, Chintadó. **South Coast:** Saija, Satinga, Saquianga, Naya, Cajambre, south of Buenaventura. **Lower Baudó:** Catrú, Dubasa, coastal rivers, Purricha, Pavaja. **Upper San Juan:** Chamí, Tadó, upper Andágueda River, southwest of Antioquia Department, Garrapatas River (north of Valle Department). **Antioquia/Córdoba:** Dabeiba, Murri, Riosucio, upper Sinú and San Jorge Rivers. **Atrato:** upper Atrato River, Capá, Bojayá, upper Baudó River, Panamá. Actually, the difference among the diverse phonological inventories is in the plosive systems and the voicing of the sibilant /s/ and the palatal affricate. Hence, the global scheme outlined in **Table 3** can be suggested.

Table 5 Grammatical similarities and differences

| | <i>Waunana</i> | <i>South Coast</i> | <i>Upper San Juan</i> | <i>Antioquia Córdoba</i> | <i>Atrato Panamá</i> |
|-----------------------------------|--------------------|--------------------|-----------------------|------------------------------|--------------------------|
| Ergative/Instrumental/Attributive | a/au/iu | a/pa | a/ba | a/ba ± ra | a/ba |
| Intransitive/Accusative | o/ta | o/ta | o/ra | o/ra/a | o/ra/da |
| Previous Reference | o | o | ra | ra | ra |
| Dative | ik | ma/ja/a | a | a | a |
| Benefactive | it ^h ee | it ^h e | ita | ita | ita |
| Sociative | dui | ome | ome | umé | ume |
| Situative | e | de | ḃe | ḃe | ḃe |
| Alative | g | ma | m/ḃa | eḃa | ḃa |
| Ablative | mu | depa | ḃeba | ḃeba | ḃeba |

Table 6 Basic sentence order

| | S | O | V |
|----------------------|--------------------------|---------------------------------|--------------------------------|
| 1. Waunana | saak-iu perro- erg | berúć tatabro | k ^h aahim mordiô |
| 2. South Coast | usa-pa perro-erg | et ^h érre gallina | peehí matô |
| 4. Upper San Juan | ûērā-ba mujer-erg | mū-buda mi pelo | kōsī cortô |
| 5. Antioquia/Córdoba | ûērā-ba mujer-erg | ûāûā-ra niño-ac | ubeasia pegô |
| 6. Atrato/Panamá | hāībana-ba chamán-erg | kauzake-da niñita-ac | uratusia frotô |

Common characteristics: 1. Predominant suffixing. 2. Occasional prefixing: integration in nominals, some verbal aspectualizing. 3. Variants of: number, gender, affection, position, permanence in the auxiliary. 4. Tactical order variation for focalization. 5. Verbalized lexical determination (adjectival verbs). 6. Actancy, opposition: agent, attributive, instrumental versus intransitive subject, accusative. 7. Great variation in prenominal suffixing. 8. Basic S O V order.

Spanish in different Embera dialects, problems of orality versus writing, and other aspects of Embera language and culture.

Bibliography

- Adam M (1888). 'Bibliographie des récentes conquêtes de la linguistique Sud-Américaine.' In *Congres internacional des Americanistes, Berlín, 7th session*. 497–520.
- Aguirre D (1990). El sintagma nominal en embera chamí de Cristianía. Master's thesis, Universidad de los Andes.
- Aguirre D (1992). 'Previsibilidad del acento en embera chamí.' In *Lenguas aborígenes de Colombia, memorias 2*. Bogotá: CCELA – Universidad de los Andes. 31–62.
- Aguirre D (1993). 'Lenguas sobrevivientes del Pacífico colombiano.' In *Pacífico, vol. 1: Colombia*. Ed. Plegra Bogotá: FEN – Universidad Nacional de Colombia. 310–325.
- Aguirre D (1994). "'parúde bú" o "Fiesta de la pubertad" entre los embera: investigación-acción participante en etnolingüística.' *Boletín de Antropología* 8(24), 43–64.
- Aguirre D (1995a). 'Fonología del embera-chamí de Cristianía (departamento de Antioquia).' In *Estudios fonológicos del grupo Choco, vol. 8: Lenguas aborígenes de Colombia*. Bogotá: Universidad de Los Andes – CCELA. 9–86.
- Aguirre D (1995b). 'Recuperación cultural y problemas prácticos de la traducción: al rescate de la tradición cultural entre los embera-chamí de Antioquia.' In *Lenguas Aborígenes de Colombia, memorias 3: La recuperación de lenguas nativas como búsqueda de identidad étnica symposium*. 7th congress of anthropology, Medellín, Colombia, 1994. Bogotá: Universidad de Los Andes – CCELA. 19–38.
- Aguirre D (1998a). 'Cosmos, naturaleza y cuerpo humano entre los embera.' In *Lenguas aborígenes de Colombia,*

memorias 5: El léxico del cuerpo humano a través de la gramática y la semántica. Bogotá: Universidad de Los Andes – CCELA. 113–120.

- Aguirre D (1998b). 'Experiencias etnoeducativas con los Embera-Chamí.' In *Lenguas aborígenes de Colombia, memorias 4: Educación endógena frente a educación formal*. Bogotá: Universidad de Los Andes – CCELA. 143–155.
- Aguirre D (1998c). 'Fundamentos morfosintácticos para una gramática embera.' In *Lenguas aborígenes de Colombia, descripciones 12*. Bogotá: Universidad de Los Andes – CCELA.
- Aguirre D (1999a). 'Ergatividad, focalización, tematización y topicalización en embera.' In *Lenguas aborígenes de Colombia, memorias 6: Congreso de lingüística Amerindia y Criolla*. Bogotá: Universidad de Los Andes – CCELA. 317–330.
- Aguirre D (1999b). *La lengua embera*. Languages of the World materials. LINCOM's Descriptive Grammar Series, no. 208. Munich: Alemania.
- Aguirre D (1999c). 'Textos y comentarios sobre vocabularios embera-chamí y embera-catio.' In *Documentos sobre lenguas aborígenes de Colombia, vol. 4: Lenguas del occidente de Colombia* [Edns. Uniandes, Colciencias.] Del archivo de Rivet P & Landaburu P (comp.). Bogotá: Universidad de Los Andes – CCELA. 83–114.
- Aguirre D, Romero F, Duque A & Gallego J (2000). *Oralidad y escritura entre los embera-chamí de Risaralda*. Pereira: Universidad Tecnológica de Pereira.
- Anonymous (1918). 'Catecismo Catío-Español para uso de los misioneros de María Inmaculada y Santa Catalina de Siena y sus neófitos catecúmenos.' In *Revista Departamental de Instrucción Pública* (2nd edn., no. 16). Medellín: Secretaría de Instrucción Pública de Antioquia. 494–513.
- Bastian A (1876). 'Bericht ubre die Spreche welche die Chamies–Andáguedas–Murindoes–Cañas Gordas–Rioverdes–Caramantas–Tadocitos–Patoes–Curasambas–Indianer Sprechen.' *Zeitschrift Für Ethnologie* 8, 359–377.
- Bedoya O L (1997). 'Los nombres compuestos en la lengua Epera de Oriente y su motivación cultural.' *Revista de Ciencias Humanas* 4(14), 86–94.
- Bedoya O L & Restrepo M (1998). *Interferencias lingüística entre la lengua y el español hablado en el choco*. Montería: Universidad de Córdoba.
- Bedoya O L & Zuluaga V (1996). 'El discurso oral y escrito entre los embera-chamí.' *Revista de Ciencias Humanas* 3(8), 45–51.
- Bollaert W (1860). *Antiquarian, ethnological and other researches in New Granada, Ecuador, Perú and Chile*. Londres: Trübner. 65–67.
- Brinton D (1891). *The American race*. New York. 175–177, 275.
- Brinton D (1894). 'Some words from the Andágueda dialect of the Choco stock.' In *Proceedings of the Philosophical Society*, vol. 34. Philadelphia. 401–402.
- Brinton D (1895). 'Vocabulary of the Noanamá dialect of the Choco stock.' In *Proceedings of the Philosophical Society*, vol. 35. Philadelphia. 202–204.

- Casas B delas (1951). *Historia de las Indias* (vols 3). México: Fondo de Cultura Económica.
- Casas B delas (1973). 'Núñez de Balboa descubre el Mar Pacífico.' In *Historiadores de las Indias*. México: Los Clásicos; New York and Panama City: W. M. Jackson.
- Casas B delas (1977). *Brevísima relación de la destrucción de las Indias*. La Habana: Ed. Ciencias Sociales.
- Castellanos J de (1942). *Historia de la gobernación de Antioquia y la del Choco*. Medellín: Biblioteca Popular de Cultura Colombiana.
- Caudmont J (1955). 'La lengua Chamí.' *Revista Colombiana de Antropología* 4, 273–293.
- Caudmont J (1956). 'La lengua Chamí' (part 2). *Revista Colombiana de Antropología* 5, 53–108.
- Chamberlain A (1907). 'South American Linguistics stocks.' *Congress international des Americanistes, 15th session*, vol. 2, Quebec. 187–204.
- Chamberlain A (1913). 'Linguistic stocks of South American Indians, with distribution map.' *American Anthropologist* 15.
- Cieza de León P de (1924). *La crónica general del Perú* (vol. 7). Lima: Colección Urteaga.
- Collins F (1879). *Vocabulary of the language of the Indians of the Cantón of Choco, State of Cauca, United States of Colombia*. [Napiquí Expedition.] Washington, DC: 118–121.
- Correa L (1981). *Breve catecismo Español y Embera Baudó*. Medellín: Cadavid Restrepo Print.
- Correa L (1982). *Cartilla Embera Baudó*. Medellín: Cadavid Restrepo Print.
- Cullen E (1851). 'Vocabulary of the language of the Cholo of Choco Indians of the Isthmus of Darién.' *Journal of the Royal Geographic Society* 20, 189–190.
- Cullen E (1866). 'The Darien Indians.' *Transactions of the Ethnological Society of London* 4, 264–268.
- Cullen E (1875). 'Description of the Choco Indians.' *New York Herald Tribune*, 167–171.
- Embera Waunana Regional Organization (OREWA) (1987). *Experiencia Educativa José Melanio Tunay. Indian Teachers Course*. Alto Buey. Quibdó: OREWA.
- Etiene C P (1887). 'Nouvelle-Grenade.' In *Aperçu Général sur la Colombia: recits de voyages en Amerique*. Gêneve. 39–41.
- Gralow F (1976). *Sistemas fonológicos de idiomas Colombianos, vol. 3: Fonología del Chamí*. Lomalinda and Meta. Instituto Lingüístico de Verano. Editorial Townsed. 29–44.
- Greenberg J (1960). 'The general classification of Central and South American languages.' In Wallace A (ed.) *Men and cultures: selected papers of the 5th international congress of anthropological and ethnographic sciences*. Philadelphia: University of Pennsylvania Press. 791–794.
- Greiffenstein C (1878). 'Vokabular der indier des Chamí.' *Zeitschrift Für Ethnologie* 10, 135–138.
- Harms P (1985). *Ne-animalará: animales en epena pedee (Saija), cartilla ILV*. Lomalinda and Meta: Ed. Townsend.
- Harms P (1987). *Typological grammar of epena pedee (Saija)*. Typewritten.
- Harms P & Powel J (1984). *Sistemas Fonológicos de idiomas colombianos, vol. 5: Fonología del epena pedee del río Saija ILV*. Lomalinda, Meta: Ed. Townsed. 155–201.
- Holmer N (1963). *Gramática comparada de un dialecto del Choco con texto, índice y vocabulario*. *Etnologiskka Studier* 26. Etnografiska Museet: Göteborg.
- Horton G (undated). Trabajo sobre los Catíos del Alto Sinú. Handwritten. Tierralta and Córdoba: Ethnographic and Linguistic Study.
- Horton G (1964). *Libro I de Epera Pedea*. Córdoba: Tierralta.
- Horton G (1965). *¡Tama! Culebra en Epera Pedea y Español*. Medellín: Tipografía Unión.
- Hoyos M (1987a). *Informe sobre la lengua embera para el atlas etnolingüístico de Colombia*. Bogotá: Instituto Caro y Cuervo.
- Hoyos M (1987b). Morfología de la palabra verbal del embera-napiquí. Master's thes., Universidad de los Andes.
- Hoyos M (1991). Rudimentos de lingüística, pedagogía y didáctica para el maestro indígena y para la enseñanza de la lectoescritura en el marco de la educación bilingüe. In Ethnoeducational Program. Plan de Universalización Básica Primaria. Bogota: Ministerio de Educación Nacional.
- Hoyos M (2000). 'Informe sobre la lengua emberá del río Napiquí.' In González M S & Rodríguez M L (comps.) *Lenguas indígenas de Colombia: una visión descriptiva*. Bogotá: Instituto Caro y Cuervo. 75–83.
- Isacson S (1973). 'Indios Cimarrones del Choco (Colombia).' *Etnografiska Museet Göteborg Arstrick*. Göteborg.
- Isacson S (1974). 'An enigmatic colonization in the XVII century in Colombia.' *Etnografiska Museet Göteborg Arstrick*. Göteborg: Göteborg Etnografiska Museum.
- Isacson S (1975). *Indiana, vol 3: Biografía Atrateña*. Berlín: Ibero-Amerikanisches Institut. Preubischer Kulturbesitz.
- Isacson S (1976). 'Embera: territorio y régimen agrario de una tribu Selvática bajo la Dominación Española.' In Friedemann N S (ed.) *Tierra, tradición y poder en Colombia*, vol. 12. Bogotá: Biblioteca Básica Colombiana.
- Lehmann W (1910). 'Ergebnisse Einer Forschungsreise in Mittelamerika und México.' *Zeitschrift für Ethnologie* 42, 695.
- Lehmann W (1920). 'Zentral Amerika.' *Zeitschrift für Ethnologie*, 1, 69–142.
- Llerena G E (2001). La predicación de la oración simple en la lengua embera del Alto Sinú. Master's thes., Universidad de Los Andes.
- Llerena G E & Llerena R (2003). *Diccionario etnolingüístico de la lengua embera*. Written for the Normal Superior de Montería. Córdoba: Montería.
- Llerena V R (1987). 'Relación y determinación en el predicado de la lengua Kuna.' In *Lenguas Aborígenes de Colombia, descripciones no. 1*. Centro Colombiano de Estudios de Lenguas Aborígenes – CCELA. Bogotá: Universidad de los Andes.
- Llerena V R (1992a). 'Estructura y variación en las fonologías de las lenguas eperá de Oriente y Occidente.' In *Lenguas Aborígenes de Colombia, memorias no. 2*.

- Centro Colombiano de Estudios de Lenguas Aborígenes – CCELA. Bogotá: Universidad de los Andes.
- Llerena V R (1992b). 'Guía de materiales para la enseñanza de la lectoescritura en la lengua epera del Alto Andágueda.' Organización Regional Embera Waunana (OREWA) CIP. Quibdó: Universidad de Antioquia.
- Llerena V R (2000). 'Elementos de gramática y fonología de la lengua Cuna.' González M S & Rodríguez M L (comps.) In *Lenguas indígenas de Colombia: una visión descriptiva*. Bogotá: Instituto Caro y Cuervo.
- Llerena V R & Gallego H (1989). 'Fonología y morfofonología de las lenguas epera-chamí de Cristianía, eperá'Catío de Jaidukama, y Kuna de Caimán Nuevo.' Centro de Documentación. Facultad de Comunicaciones. Medellín: Universidad de Antioquia.
- Llerena V R & Gallego H (1990). 'Resultados de la investigación fonológica de las lenguas Amerindias de Antioquia.' *Cultura Embera*. Memories of the 4th congress of anthropology in Colombia. Medellín: Organización Indígena de Antioquia (OIA). 68–90.
- Loboguerrero M (1976). *Lingüística del Chamí*. Fieldwork. Bogotá: Universidad Nacional de Colombia.
- Loewen J (1954). *Waunana grammar: a descriptive analysis*. Thes., University of Washington.
- Loewen J (1958). An introduction to Epera speech: Sambu dialect. Ph.D. diss., University of Michigan.
- Loewen J (1960). 'Dialectología de la familia lingüística Choco.' *Revista Colombiana de Antropología* 9, 8–22.
- Loewen J (1963). 'Choco: phonological problems.' *International Journal of American Linguistics* 29, 357–371.
- Loewen J (1964). 'The Choco and their Spirit World.' *Practical Anthropology* 11(3), 97–104.
- Loewen J (1972). 'El cambio cultural entre los Choco de Panamá.' *América Indígena* 32(1).
- Lotero L (1972). *Monografía de los Indígenas Noanamá*. Medellín: Editorial Servigráficas.
- Loukotka C (1968 [1942]). *Classification of South American Indian languages*. Los Angeles: University of California.
- Manzini G (1973). *Abecedario Embera del alto río San Juan*. Vicariato Apostólico de Itsmina. Medellín: Yepes Print.
- Manzini G (1974). *Indígenas e indigenismo en el Choco*. Medellín: Editorial Universidad de Antioquia.
- Manzini G (1976). 'El más antiguo testimonio sobre el lenguaje emberá.' *Revista de la Universidad del Choco* 1, 101–108.
- Martínez E & Guisao E (1980). *Mi cartilla Catía: texto de lectura bilingüe Katío-Español*. Secretaría de Educación y Cultura de Antioquia. Medellín: Talleres Albón.
- Mason A (1950). *Handbook of South American Indians*, vol. 6: *The languages of South American Indians*. Washington, D.C: Bureau of American Ethnology. 157–317.
- Meillet A & Cohen M (1952). *Les langues du monde*. Paris: Centre National de la Recherche Scientifique, 1152–1160.
- Mejía G (1987). Relaciones actanciales y de persona en Waunana. Master's thes., Universidad de los Andes.
- Mejía G (2000a). 'Lenguas aborígenes de la costa pacífica de Colombia.' In González M S & Rodríguez M L (comps.) *Lenguas indígenas de Colombia: una visión descriptiva*. Bogotá: Instituto Caro y Cuervo.
- Mejía G (2000b). 'Presentación y descripción fonológica y morfosintáctica del waunana.' In González, M S & Rodríguez M L (comps.) *Lenguas indígenas de Colombia: una visión descriptiva*. Bogotá: Instituto Caro y Cuervo.
- Michael R & Nellis M J (1984). *¿Na Karema? El Abecedario ebera (chamí), cartilla 1*. Lomalinda and Meta: L. V. Editorial Townsend.
- Mollien G (1824). *Travels in the Republic of Colombia in the years 1822 and 1823*. London. ([1944]. 'Viaje por la República de Colombia.' *Biblioteca Popular de Cultura Indiana*, vol. 8. Bogotá.)
- Mollien G (1979). 'Recorrido por la Tierra del Oro.' In *Las Maravillas de Colombia*, vol. 1. Bogotá: Editorial Forja. 113–134.
- Nordenskiöld E (1927). 'The Choco Indians of Colombia and Panama.' *Discoverg* 8(95), 350–377.
- Nordenskiöld E (1928). 'Les Indiens del' Isthmus de Panamá.' In *La Geographie*. París. 229–319.
- Ortega R C (1978). *Los estudios sobre lenguas indígenas en Colombia*. Bogotá: Imprenta del Instituto Caro y Cuervo.
- Ortiz S E (1937). 'Clasificación de las lenguas indígenas de Colombia.' *Idearium* 1, 76.
- Ortiz S E (1940). 'Lingüística Colombiana: familia Choco.' *Universidad Católica Bolivariana* 6, 46–77.
- Ortiz S E (1954). *Estudio sobre lingüística aborígen de Colombia*. Bogotá: Ed. Kelly. 271–310.
- Ortiz S E (1965). *Historia extensa de Colombia*, vol. 1, book 3: *Lenguas y dialectos indígenas de Colombia*. Bogotá: Ediciones Lerner.
- Pardo M (1981). 'Bibliografía sobre indígenas Choco.' *Revista Colombiana de Antropología* 23, 464–528.
- Pardo M (1983). 'Transformaciones históricas en los indígenas Choco.' *Boletín de Antropología* 5, 611–628.
- Pardo M (1984). Etnolingüística de indígenas Choco. Investigation report presented to the Fondo para la Promoción de la Investigación y la Técnica (FPIT) del Banco de la República, Bogotá.
- Pardo M (1985a). 'Las lenguas Choco en Antioquia: aspectos fonológicos.' *Revista de Ciencias Sociales* 1, 53–68.
- Pardo M (1985b). 'Los indígenas chocoanos, 450 años de resistencia.' *Revista Codechocó* 1, 45–50.
- Pardo M (1986). 'La situación de la lingüística sobre el grupo Choco.' In *Memorias de eventos científicos Colombianos*, no. 42: *Memorias III congreso de antropología en Colombia*. Instituto Colombiano para el Fomento de la Educación Superior. Bogotá. 87–117.
- Pardo M (1987). 'Regionalización de indígenas Choco.' In *Revista del Museo del Oro Boletín* 18, January–April. Bogata: Museo del Oro. 46–63.
- Pardo M (1989). 'Lengua y sociedad: indígenas Choco.' *Arqueología: Revista de los Estudiantes de Antropología de la Universidad Nacional* 10, 69–74.
- Pardo M (1997). 'Aspectos sociales de las lenguas Choco.' In Pachón X & Correa F (comps.) *Lenguas Amerindias: condiciones sociolingüísticas en Colombia*. Bogotá: Instituto Caro y Cuervo–Instituto Colombiano de Antropología (ICAN). 321–381.
- Pardo M & Aguirre D (1993). 'Dialectología Choco.' In Rodríguez M L (comp.) *Estado actual de la clasificación*

- de las lenguas indígenas de Colombia. Colección Ezequiel Uricoechea, no. 11. Bogotá: Instituto Caro y Cuervo. 269–312.
- Pardo M & Orozco L A (1987). Cartilla 1: *ebera be'dea. karta jarabada 'buubada*. Cartilla 2: *ebera be'dea. oar-rará karta jaradiábadaa*. Codechoco-Diar-Orewa.
- Peláez T (1885). 'Vocabulario de los Indios del Río Verde, Mutatá, Dabeiba, Frontino, Cañasgordas.' In Uribe M (ed.) *Geografía general y compendio histórico del Estado de Antioquia en Colombia*. París: Goupy et Jourdan Press. 535–541.
- Picón M L (1985). *Maach meua waunana esap gaai k'aug-tarrau*. Cartilla Vicariato Apostólico de Itzmina. Quibdó: Educación Contratada Quebrada de Pichimá.
- Pinart A (1887). 'Les Indiens de l'état de Panamá.' In *Revue D'Etnographie*, vol. 6. Paris. 33–56.
- Pinart A (1897). 'Vocabulario Castellano-Choco' (Baudó-Citaræ). *Petite Bibliotheque Americaine* vol. V.
- Pinto C (1950). *Diccionario Catío-Español y Español-Catío*. Manizales: Imprenta Departamental.
- Pinto C (1974). *La cultura Catía: su lengua y su cultura*, vol. 2: *La lengua: gramática y diccionario*. Medellín: Editorial Granamérica.
- Pinto C (1977). 'Los Indios Chamí.' *Aleph*, 21–35.
- Pinto C (1978). *La cultura Catía: su lengua y su cultura*, vol. 1: *Culturá y mithología*. Medellín: Ed. Compás.
- Prado M (1982). El epera de Saija: un estudio inicial. Master's thes., Universidad del Valle.
- Prado M (1985). *Epera pedee páda: lengua nativa escrita*. Vocabulary and Alphabet. Cartillas de Lecto-escritura. Popayán: Universidad del Cauca.
- Prado M (1990). 'La decisión de escribir una lengua y sus implicaciones: el epera de Saija, un caso.' In *Cultura Embera, memories of the 4th congress of anthropology in Colombia*. Medellín: Organización Indígena de Antioquia (OIA). 58–67.
- Prado M (1991). Aproximaciones a la nasalidad del epera de Saija. Speech presented to the Segundo Congreso del CCELA, Villeta, Cundinamarca.
- Prado M (1992). *Problemas lingüísticos y conflictos inter-étnicos entre comunidades negras e indígenas epera de Saija*. Cali: Universidad del Valle.
- Rasmussen E (1985). *Nociones gramaticales de Eberá*. Panamá: Universidad Nacional de Panamá.
- Rasmussen E (1986). *Principales rasgos fonológicos del Embera*. Arosemena M (trans.). Panamá: Universidad Nacional de Panamá.
- Reichel-Dolmatoff G (1945). 'Bibliografía lingüística del grupo Choco.' *Boletín de Arqueología* 1(6).
- Reichel-Dolmatoff G (1955). *Diario de viaje por las antiguas provincias de Cartagena del Padre Joseph Palacios de la Vega*. Bogotá: Ministerio de Educación.
- Reichel-Dolmatoff G (1964). *Bibliografía del Choco*. Universities of Cali and Tulane.
- Rex E (1975). On Catío grammar. Thes., Universidad de Texas at Arlington.
- Rex E & Schotlendreyer M (1973). 'Sistema fonológico del Catío.' In *Sistemas fonológicos Colombianos*, vol. 2. Lomalinda and Meta: Ed. Townsed. Instituto Lingüístico de Verano. 73–88.
- Rivet P (1912). 'Les familles linguistiques du nor-ouest de L'Amérique du Sud.' *L'Anne Linguistique* 4, 123–126.
- Rivet P (1924). 'Langues de L'Amérique du sud et des Antilles.' In Meillet et al. (eds.) *Les langues du monde*, vol. 16. Paris. 639–712.
- Rivet P (1943). 'La Influencia Karib en Colombia.' *Revista del Instituto Etnológico Nacional* 1, 13–96.
- Rivet P (1944). 'La Lengua Choco.' *Revista del Instituto Etnológico Nacional* 1, 297–349.
- Rivet P (1946). 'Groupe Catío.' *Linguistique: Journal de la Société des Americanistes* 34, 25–29.
- Robledo E (1922). 'Vocabulario de los Chamíes.' *Repertorio Histórico de Antioquia* 4(5–8), 603–607.
- Romero Loaiza F & Bedoya O L (1997). 'La enseñanza del español como segunda lengua en los embera-chamí y nasa, una propuesta lingüística y pedagógica.' *Boletín de Antropología* 11(29), 11–19.
- Romoli K (1961). Pascual de Andagoya y el descubrimiento de la Costa Pacífica. Speech presented to Congreso Internacional de Historia Hispanoamericana, Cartagena.
- Romoli K (1962). 'El sureste del Cauca y sus indios al tiempo de la Conquista Española, según documentos contemporáneos del Distrito de Almaguer.' *Revista Colombiana de Antropología* 11, 239–297.
- Romoli K (1975). 'El Alto Choco en el siglo XVI.' *Revista Colombiana de Antropología* 19, 9–38.
- Romoli K (1976). 'El Alto Choco en el siglo XVI (pt.2): 'las gentes.' *Revista Colombiana de Antropología* 20, 25–78.
- Sánchez M & Castro C (1977). 'Una gramática pedagógica del Waunana' (pt. 1). In *Lenguas de Panamá*, vol. 3. Panamá: Instituto Nacional de Cultura. I. L. V.
- Santísimo Sacramento F P del (1936). El idioma Katío. Essay. Medellín.
- Schotlendreyer M (1973). *El abecedario ebena (Catío) ¿Cane busia Cobúa?* Lomalinda and Meta: I. L. V. Editorial Townsed.
- Schotlendreyer M (1977). 'Estudios en Camsá y Catío: la narración folclórica catía como un drama en actos y escenas.' In *Serie sintáctica*. Lomalinda and Meta: I. L. V. Editorial Townsed. 97–153.
- Seeman B (1851). 'The aborigens of the Isthmus of Panamá.' *Transactions of the American Ethnological Society* 3, 179–181.
- Simón F P (1882). *Noticias historiales de las Conquistas de tierra firme en las Indias occidentales* (5 vols). Bogotá: Medardo Rivas Press.
- Simons F A (1887). 'Vokabular des Tukurá.' *Zeitschrift für Ethnologie* 19, 302.
- Stansell D (1973). 'Embera.' In *Aspectos de la cultura material de grupos étnicos de Colombia*, vol. 1. Lomalinda and Meta: I. L. V. Editorial Townsed. 179–194.
- Torres de Arauz R (1960). 'Los grupos humanos del Darién Panameño.' Report and annexes of subcommittees of the Darién, 8th congreso Panamericano, Carreteras, Panamá.
- Torres de Arauz R (1966). *La cultura Choco: estudio etnológico e histórico*. Panamá: Universidad de Panamá.
- Torres de Arauz R (1971). 'Culturas Prehispánicas del Darién.' *Hombre y Cultura* 2(2), 7–40.

- Torres de Arauz R (1972). 'Panorama actual de las culturas indígenas panameñas.' *América Indígena* 32(1), 77–94.
- Tovar A (1961). *Catálogo de lenguas de América del Sur*. Buenos Aires: Ediciones Suramericana.
- Tovar A & Larrucea C (1984). *Catálogo de las lenguas de América del Sur*. Madrid: Editorial Gredos.
- Uribe J V (1881). 'Gramática y vocabulario de la lengua que hablan los indios darienes que habitan la región comprendida entre las desembocaduras del Atrato, en el Atlántico, y del San Juan, en el Pacífico, y la cordillera en que limitan las antiguas provincias del Choco y Antioquia.' *Congreso internacional de Americanistas*, session 4, vol. 2. Madrid. 297–309.
- Uribe J V (1883). 'Gramática y vocabulario de la lengua que hablan los indios darienes.' *Congreso internacional de Americanistas*, session 4, vol. 2. Madrid. 269–309.
- Vargas P (1984). La Conquista tardía de un territorio aurífero: la reacción de los Embera de la cuenca del Atrato a la Conquista Española. Thes., Universidad de los Andes.
- Vargas P (1986). Los habitantes del Alto Río Sinú y sus fronteras, siglos XVI–XVIII. Report to Instituto Colombiano de Antropología (ICAN), Bogotá.
- Vargas P (1990). Los Embera y los Kuna: impacto y reacción ante la ocupación española, siglos XVI y XVII. Master's thes., Universidad del Valle.
- Velásquez R (1916). 'Vocabulario de los Chamés.' *Boletín de la Sociedad de Ciencias Naturales del Instituto de la Salle* 4, 147–150.
- Wassen H (1935). 'Notes on the southern groups of Choco Indians in Colombia' (Waunana and Saija). *Etnologiska Studier* 1, 35–182.
- White R B (1884). 'Notes on the aboriginal races of the north western provinces of South America.' *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 13, 240–456.

Chorasmian

P O Skjærvø, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Chorasmian was an Iranian language spoken in medieval Chorasmia, a state on the Oxus/Amu Darja south of the Aral Sea. The name is first mentioned in the Avesta and the Achaemenid inscriptions (see **Avestan; Persian, Old**) but the language is known only from much later times. Several words pertaining to the calendar and astronomy were cited by Abu Rayhan Biruni in his *Athar al-baqiya* (comp. 1000). Since then archeological excavations have uncovered inscriptions and documents on parchment and wood from ca. 200–700 A.D.; also, a number of manuscripts of Arabic works containing interlinear glosses in Chorasmian have been found in libraries in Turkey, notably Abu'l-Qasim Zamakhshari's *Muqaddimat al-adab* (ms. from ca. 1200) and several 13th-century Arabic law books. The Chorasmian glosses are written in Arabic script, with several modified letters. Those in the *Muqaddima* are often underpointed or not pointed at all, which makes them hard to interpret.

Some Arabo-Persian letters were modified to express special Chorasmian sounds. Triple superscript dots over $c = ts$ and dz , over $f = \beta$. Triple subscript dots were used under s to indicate s , not \check{s} , and single subscript dot under d to indicate d , not δ .

Chorasmian historical phonology is characterized by extensive affrication of dentals, palatalization, and

a variety of, often unpredictable, simplifications of consonant groups. For instance, t and $d > c$ [ts] and j [dz] before and after i , y : $pc < pati-$ (preverb) and $*pitā$ 'father'; pzy 'sinew', cf. Av. $paidiā-$. Intervocalic \check{s} developed variously: $'mh$ 'ewe', cf. Av. $maēš-$; mwf 'mouse', cf. Av. $mūš$; ywx 'ear', cf. MPers. $gōš$; etc.

The Chorasmian vowel system is characterized by the reappearance (in the script) of final vowels before suffixes, $pc = pic^a$ 'father', but $pc'm = picā-m^i$ 'my father'. Contraction of final vowels with vowels of suffixes is common, e.g., $hāβir-m^a$ 'give.IMPERF-1ST.SING' = 'I gave', but $hāβir-n^a-h^i-d^i$ 'give.IMPERF-1ST.SING-he/she/it. ENCL.OBL-you.ENCL.OBL' = 'I gave her to you' $> hāβir-nā-h^i-d^i > hāβir-n-i-d^i$. Such final vowels are sometimes indicated by the Arabic vowel marks.

Masculine and feminine gender are distinguished in the definite article (\check{i} , ya ; $-i$, $-ā$ after prepositions) and in declension (nom. sing. masc. no ending, but fem. $-a$). Five cases are distinguished in masculine nouns: nominative-accusative, vocative ($-a$), possessive ($-ān$), dative ($-i$), and ablative-locative ($-a$), but feminine nouns have only two forms: nominative and locative ($-a$) contrasting with the other cases ($-iya$). The plural endings are $-i$ or $-ina$, possessive $-in-ān$. A final $-k$ becomes $-c$ before $-i$. The direct object can be marked by $-dār$ attached to the dative (presumably $< *rād$, cf. Pers. $-rā$). Examples: $\bar{i} kām-h^i$ 'DEF mouth.MASC-he. ENCL.OBL' 'his mouth'; $f-\bar{i} kāmā-h^i$ 'in DEF mouth'; $yā camā-h^i$ 'DEF eye.FEM-he.ENCL.OBL'; $yā cam-yā-h^i dār$ 'DEF eye.FEM-DAT-he.ENCL.OBL' = 'his eye' (DO); $\bar{i} βandik$ 'the servant', $\bar{i} bandic-i$ 'the servants', $f-\bar{i} bandic-i-h^i$ 'with-DEF servant-PL-he.ENCL.OBL' = 'with

his servants'; *ī bfin-ēnik ī βūm-in-ān* 'DEF create.AGT DEF earth-PL-POSS' = 'the creator of the earths'.

When several enclitic personal and local pronouns are added to a verb, the order is strict, e.g., *γēr-idā-hī-nā-bir* 'turn-IMPERF.3RD.SING-he.ENCL.OBL-they.ENCL.OBL-upon' = 'he made them go around him', where *-bir* goes not with the preceding *-nā-*, but with *-hī-*. When personal and local complements follow the verb, they must be anticipated as enclitics, e.g., *m-uxwās-idā-nā-w^a f-ī razik-^a ī cūb* 'IMPERF-let-PAST.3RD.SING-they.ENCL.OBL-there in-DEF-vineyard-LOC DEF-water.PL' = 'he let the water into the vineyard'; *hīδ-idā-hī-nā-dā-bir ī salām* 'read-IMPERF.3RD.SING-it.ENCL.OBL-they.ENCL.OBL-there-on DEF-greeting.PL' = 'he recited the greetings upon him'.

The verbal system is of the Eastern Middle Iranian type. There are three stems: present, past, and perfect (perfect participle = past stem + suffix *-ik*, FEM *-ic^a*). There are numerous modal forms (indicative, imperative, subjunctive, optative, injunctive); an imperfect formed with prefixes (*m-ikk-* 'did') or lengthening of the vowel of the first syllable (*h-ā-βir-* 'gave'), both reflexes of the Old Iranian augment; a form ending in *-i(n)* added to personal endings, the function of which is not completely clear but which is referred

to as 'permansive'; a (present) perfect formed with the perfect participle and the verb *δār-* (transitive verbs) or 'be' (intransitive verbs), e.g., *akt-ik δāriy-ā-yī* 'do.PERF.PART.MASC have.PRES-1ST.SING-PERMAN SIVE' = 'I may have done'; *purāc^a-ihi* [*<purād-c-⟩*] *purācīhi* 'divorce.PERF.PART.FEM-be.PRES.2ND.SING' = 'you are divorced'.

Bibliography

- Benzing J (1968). *Das Chwaresmische Sprachmaterial einer Handschrift der »Muqaddimat al-adab« von Zamaxšari I: Text*. Wiesbaden: Steiner.
- Benzing J (1983). *Chwaresmischer Wortindex*. Taraf Z (ed.). Wiesbaden: Harrassowitz.
- Henning W B (1971). *A fragment of a Khwarezmian dictionary*. In MacKenzie D N (ed.). London: Lund Humphries.
- Humbach H (1989). 'Choresmian.' In Schmitt R (ed.) *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert. 193–203.
- MacKenzie D N (1990). *The Khwarezmian element in the Qunyat al-Munya*. Amarat H & MacKenzie D N (trans.). London: School of Oriental and African Studies.
- Samadi M (1986). *Das Chwaresmische Verbum*. Wiesbaden: Harrassowitz.

Chukotko-Kamchatkan Languages

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

Chukotko-Kamchatkan

Chukotko-Kamchatkan, formerly also known as Luor[a]vetlan, is a small family of languages spoken in extreme northeastern Siberia on the Chukotka Peninsula, opposite Alaska and the large Kamchatka Peninsula in far Eastern Siberia. The family consists of four remaining languages, Alutor, Chukchi, Itelmen, and Koryak. All of the languages in the group, excluding Chukchi, are endangered; Kerek became extinct in the 1990s.

Alutor (*Ethnologue* code ALR), also known as Alyutor or Palana Koryak, is spoken by some 200 people in the villages of Vyvenka and Rekiniki in the Koryak National District, in the northeast Kamchatka Peninsula. Chukchi (*Ethnologue* code CKT) is spoken by some 10 000 people, primarily on the Chukchi Peninsula of northeastern Siberia. In English language literature, especially older works,

the language is sometimes spelled Chukchee as well. Several local variants exist, but differences are relatively minor. More celebrated were the once active phonological differences in men's and women's speech, seen in the following word pair: (men) *reqər-kən* = (women) *tzeqətzən* 'what is s/he making/doing?' (Kämpfe and Volodin, 1995: 8). Itelmen (*Ethnologue* code ITL) is also known as Kamchadal. Itelmen is currently moribund, with fewer than 100 speakers. Itelmen speakers are found primarily in the Tigil region, in Kovran, and in the Upper Khairiuzovo villages on the Kamchatka Peninsula. There were originally at least three Itelmen languages, two gradually giving way to Russian over the past two centuries, and they are now extinct. Only the Western dialect remains; it is sometimes divided into separate Kovran and Sedanka varieties. Kerek (*Ethnologue* code KRK) became extinct in the late 1990s. It was closely related to Koryak (*Ethnologue* code KPY); Koryak has some 3500 speakers scattered across the Koryak National Okrug, on the northern half of Kamchatka. An alternate name is Nymylan. There are several divergent varieties, some now

considered separate languages (Alutor). Dialects include Chavchoven, Apukin, and Kamen.

Itelmen stands in isolation from the northern languages genetically, with the speech representing a southern branch. It is sometimes debated whether Itelmen is related at all to Northern Chukotko-Kamchatkan, and it is indeed different in numerous ways, but these are attributable rather to different substrate populations and various locally defined internal developments within Northern and Southern Chukotko-Kamchatkan, and their ultimate genetic unity seems clear. The northern branch in many interpretations has further subgroups of Alutor and Koryak (and Kerek), in opposition to Chukchi.

Along the coasts, Chukchi people live as sea mammal hunters, like the local Yup'ik populations, but they live as reindeer herders in the interior. Approximately three-quarters of the Chukchi live as reindeer herders. Northern Kamchatkan groups mainly practice reindeer-oriented economies and fishing and sea mammal hunting along the coasts. The Itelmen live primarily as subsistence fishers.

Chukotko-Kamchankan languages in general, but the northern ones in particular, are characterized by a range of features that set them apart from many indigenous Siberian languages, but also reflect a number of areally common features. First, many words in Chukotko-Kamchatkan languages are very long (e.g., Chukchi *ga-npənačg-ərgəna-qora-ma* 'with the old men's reindeers' (Skorik, 1986: 107)), and initial *ŋ*-is common (as is typical of northern and eastern Siberian languages (Anderson, 2003). Clusters of stop + *ŋ* are also found. Example (1) is from Skorik (1986: 79, 85) (cf. Itelmen *ŋosx* 'tail' and *ŋeyŋe* 'mountain' (Volodin, 1976: 31)):

| (1) Chukchi | Koryak | Alutor | Kerek | gloss |
|-------------|--------|--------|--------|--------------|
| ŋoyŋən | ŋoyŋən | ŋoyŋən | ŋuyŋən | 'tail' |
| ŋəron- | ŋəyon- | ŋərun- | nəyuy- | '3 together' |
| laŋ-o | laŋ | laŋ | laŋu | |

Compare Kerek *tŋivek* 'to send' (Skorik, 1986: 89) with Itelmen *pŋilpnəl* 'root' (Skorik, 1986: 78)). Itelmen shows an unusual tolerance to consonant clusters word-initially, as well as ejective consonants that the northern languages do not share. Thus, words such as *klfknan* 'it fell out' and *kstk'łknan* 'he jumped' may be found in Itelmen.

Northern Chukotko-Kamchatkan languages stand out for their areally atypical system of vowel harmony. Vowels belong to one of two harmonic classes, strong/dominant and weak/recessive. A strong vowel triggers strong allophones throughout the word, and therefore a vowel in an affix may trigger alternation in stem vowels, as shown in Example (2) for Koryak:

| | | | |
|-----------|------------|---------------|-------------------|
| (2) weyem | 'river' > | wayaməŋ | 'river-DAT' |
| mil'ut | 'hare' > | mel'otaŋ | 'hare-DAT' |
| eŋpič | 'father' > | aŋpeče-na-naŋ | 'father-AUGM-DAT' |

Note: *geyqə-miml-e* 'with water' vs. *gawən-meml-əma* 'with water' (Zhukova, 1972: 111–112; 120).

Among the most characteristic features of Chukotko-Kamchatkan morphology is the frequent use of circumfixes (combined prefix + suffix combinations) to encode a variety of inflectional categories, both nominal and verbal, some of which appear to be very old in the family. In Koryak, this is realized as *ga-čol'-ma* 'with salt' (Zhukova, 1972: 120), and in Chukchi it is *ga-npənačg-ərgəna-qora-ma* 'with the old men's reindeers' (Skorik, 1986: 107). In Koryak, *y(A)-...-ŋ* (Zhukova, 1972: 202):

| | |
|-----------------------|-----------------------|
| (3) y-ačəgaŋ-ŋ-ək | ye-lqəŋ-ŋ-ək |
| DESID-laugh-DESID-INF | DESID-leave-DESID-INF |
| 'want to laugh' | 'want to leave' |

In Chukchi, *re-...-ŋ* (Kämpfe and Volodin, 1995: 88):

| | | |
|------------------|---|---------------------|
| (4) vinretə-rkən | > | re-vinretə-ŋə-rkən |
| help-IMPERF. | | DESID-help- |
| REALIS | | DESID-IMPERF-REALIS |
| 'he helps' | | 'he wants to help' |

Among the wider relationships that have been proposed for Chukotko-Kamchatkan languages, none widely accepted by specialists, are connections with Uralic, Eskimo-Aleut, and 'Eurasian,' among others.

Bibliography

- Anderson G D S (2003). 'Towards a phonological typology of Native Siberia.' In Holisky D A & Tuite K (eds.) *Current trends in Caucasian, East European and Inner Asian linguistics. Papers in honor of Howard I. Aronson*. Amsterdam & Philadelphia: John Benjamins. 1–22.
- Angere J (1951). 'Das Verhältnis des tschuktschischen Sprachgruppe zu dem uralischen Sprachstamme.' *Språkvetenskapliga Sällskapets i Uppsala Förhandlingar* 1949–1951, 109–150.
- Asinovskij A S (1991). *Konsonantizm chukotskogo jazyka*. Leningrad: Nauka.
- Bobaljik J (1998). 'Pseudo-ergativity in Chukotko-Kamchatkan agreement systems.' *Recherches Linguistiques de Vincennes* 27, 21–44.
- Bogoraz W (1922). 'Chukchee.' In Boas F (ed.) *Handbook of American Indian languages*. Washington: Government Printing Office. 631–903.
- Bogoraz V G (1937). 'Luoravetlanskij (chukotskij) jazyk.' In Krejnovich E A & Koshkin J P (eds.) *Jazyki i pis'mennost' narodov Severa. Čast' III. Jazyki i pis'mennosti paleoaziatskix narodov*. Moscow: Gosudarstvennoe učebno-pedagogičeskoe izdatel'stvo.

- Comrie B (1979). 'Degrees of ergativity: some Chukchee evidence.' In Plank F (ed.) *Ergativity*. New York: Academic Press. 219–240.
- Comrie B (1980). 'Inverse verb forms in Siberia: evidence from Chukchee, Koryak, Kamchadal.' *Folia Linguistica* 1(1), 61–74.
- Dürr M, Kasten E & Khalojmova K N (2001). *Itelmen language and culture*. Münster/Berlin: Waxmann [multi-media CD-ROM].
- Fortescue M (1998). *Language relations across Bering Strait: reappraising the archaeological and linguistic evidence*. London: Cassel.
- Georg R-S & Volodin A P (1999). *Die itelmenische Sprache*. Wiesbaden: Harrassowitz.
- Inenlikej P I (1987). *Slovar' chukotsko-russkij i russko-chukotskij* (2nd edn.). Leningrad: Proveshchenie.
- Kämpfe H R & Volodin A P (1995). *Abriss der tschuktschischen grammatik*. Wiesbaden: Harrassowitz.
- Kibrik A E, Kodzasov S V & Muravyova I A (2000). *Aljutorskij jazyk i folklor*. Moscow: IMLI RAN Nasledie.
- Kozinsky I S, Nedjalkov V P & Polinskaja M S (1988). 'Antipassive in Chukchee: oblique object, object incorporation, zero object.' In Shibatani M (ed.) *Passive and voice*. Amsterdam: John Benjamins. 651–706.
- Krejnovich E A (1979). 'Nejtral'nyj glasnyj i konsonatizm v chukotsko-kamchatskikh jazykakh.' In *Zvukovoj stroj jazykov*. Leningrad: Nauka. 157–166.
- Mudrak O A (2000). *Etimologicheskij slovar' Chukotsko-Kamchatskix jazykov*. Moscow: Jazyki Russkoj Kul'turi.
- Nedjalkov V P (1979). 'Degrees of ergativity in Chukchee.' In Plank F (ed.) *Ergativity*. New York: Academic Press. 241–262.
- Nedjalkov V P, Inenlikej P I & Rachtilin V G (1983). 'Rezultativ i perfekt v Chukotskom jazyke.' In *Tipologija rezultativnyx konstruksij*. Leningrad: Nauka. 101–109.
- Radloff L (1861). *Über die Sprache der Tschuktschen und ihr Verhältnis zum Korjakischen*. St. Petersburg: Kommissionäre der Kaiserlichen Akademie der Wissenschaften.
- Skorik P J (1948). *Očerki po sintaksisu chukotskogo jazyka. Inkorporatsija*. Moscow: Gosudarstvennoe uchebno-pedagogicheskoe izdatel'stvo.
- Skorik P J (1961). *Grammatika chukotskogo jazyka. Chast' I*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Skorik P J (1977). *Grammatika chukotskogo jazyka. Chast' II*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Skorik P J (1986). 'Kategorii imeni sushchestvitel'nogo v Chukotsko-Kamchatskikh jazykakh.' In Skorik P J (ed.) *Paleoaziatskie jazyki*. Novosibirsk: Akademija Nauk SSSR. 76–111.
- Spencer A (1995). 'Incorporation in Chukchi.' *Language* 71(3), 439–489.
- Stebnitskij S N (1937). 'Osnovnye foneticheskie razlichija dialektov nymylanskogo (korjaksckogo) jazyka.' In *Pamjati V. G. Bogoraza*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Stebnitskij S N (1938). 'Aljutorskij dialekt nymylanskogo jazyka.' *Sovetskij Sever* 1, 65–102.
- Volodin A P (1976). *Itel'menskij jazyk*. Leningrad: Nauka.
- Volodin A P (1991). 'Prospekt opisaniya grammatiki kerekckogo jazyka (Chukotsko-Kamchatskaja gruppa).' In *Jazyki narodov sibiri. grammaticheskie issledovanija*. Novosibirsk: Nauka Sibirskoe Otdelenie.
- Volodin A P (1997). 'Itel'menskij jazyk.' In *Jazyki Mira: Paleoaziatskie jazyki*. Moscow: Indrik. 60–72.
- Volodin A P & Khalojmova K N (1989). *Itel'mensko-russkij slovar'*. Leningrad: Proveshchenie.
- Worth D S (1962). 'La place du Kamtchadal parmi les langues soi-disant paléosibériennes.' *Orbis* XI(2).
- Zhukova A N (1968). 'Aljutorskij jazyk.' In *Jazyki narodov SSSR, V, Paleoaziatskie jazyki*. Moscow: Akademija Nauk SSSR. 294–309.
- Zhukova A N (1972). *Grammatika korjaksckogo jazyka*. Leningrad: Nauka.
- Zhukova A N (1980). *Jazyk palanskikh korjakov*. Leningrad: Nauka.

Church Slavonic

C M MacRobert, Oxford University, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Church Slavonic is a generic term for the closely related, highly conservative varieties of Slavic language used for liturgical purposes by the Eastern Orthodox Slavs (Belorussian, Bulgarian, Macedonian, Russian, Serbian, Ukrainian) and the Ukrainian Uniates, and also by the Romanians until the 16th century and, until the introduction of services in the vernacular, the Roman Catholic Croats of the Slavonic rite. In the medieval period, Church Slavonic

also had the wider functions of a literary language among most of these peoples.

Church Slavonic originated in the translations of Scripture and liturgy made mainly from Greek by SS Cyril and Methodius and their associates in the late 9th and early 10th centuries (see **Old Church Slavonic**). The basic vocabulary, grammatical forms, and pronunciation of these texts predominantly followed the usage of Slavs in the southeast Balkans, while syntax and word-formation were to a large extent modeled on Greek.

Two developments signal the transition, by the end of the 11th century, from Old Church Slavonic to Church Slavonic. One was the emergence of local

varieties, such as Croatian, Russian, and Serbian Church Slavonic, which compromised between traditional pronunciation and grammatical forms and the vernacular usage of the area. Initially unsystematic, these modified varieties rapidly stabilized to local norms that in the hands of competent scribes attained a high degree of regularity. The other development consisted in revisions of syntax and vocabulary, which seem to have been motivated partly by the desire to eliminate outdated or unfamiliar linguistic material, but also aimed to make texts conform to a received Greek version and to produce more closely literal translations. The earliest systematic revisions are associated with Preslav, the capital of Bulgaria in the 10th century, when a number of early Church Slavonic revisions, new translations, and original compositions came into existence. There also appears to have been a revision of Croatian Church Slavonic texts on the basis of Latin sources in the 12th century.

Revisionist tendencies culminated by the 14th century in comprehensive reform of scriptural and liturgical translations into Bulgarian and Serbian Church Slavonic. This development has been associated with the Bulgarian patriarch Euthymius (elected patriarch in 1375; exiled by the Turks in 1393), though more recent research suggests it began in the early part of the century, perhaps on Mount Athos. The resulting standardized orthography, conservatism in grammatical forms and vocabulary, and highly literalistic translational practice were introduced among the East Slavs from the end of the 14th century, albeit with some adjustments to pre-existing local usage. The late 16th and early 17th centuries saw attempts in the Ukraine at systematic description of this late and composite type of Church Slavonic, on the model of Greek and Latin grammars; the most comprehensive of these, compiled by Meletij Smotryč'kyj in the early 17th century and subsequently modified to conform to Muscovite practice, remained the fullest description of Russian Church Slavonic until the 19th century.

Further revisions of Church Slavonic texts initiated in Muscovy or the Ukraine in the 16th and 17th centuries, though controversial in their time, dealt with minor textual discrepancies or the detail of grammatical and orthographical norms. A final standardization was effected in the publications approved by the Synod of the Russian Orthodox Church in the 18th century. Thanks to the dissemination of these printed books in the Balkans, the Orthodox Bulgarians, Macedonians, and Serbs came to use 'Synodal' Russian Church Slavonic, albeit with their own pronunciations.

Modern Church Slavonic does not stand in a simple genetic relationship to other Slavic languages. Its

texts may be understood in different ways and to varying degrees by Slavs of differing linguistic background and, as a result of literalistic translational practices aiming at morpheme for morpheme equivalence, some of them are intelligible only with the help of their Greek originals. It is virtually a closed system, for though new texts can be created if need arises, they are acceptable as Church Slavonic only insofar as they reproduce traditional constructions and phraseology. While its liturgical use still prevails in the Russian Orthodox Church, among the Orthodox South Slavs, Church Slavonic tends increasingly to be supplanted by modern vernacular translations, and survives mainly as a vehicle for the traditional corpus of hymns.

Bibliography

- Corin A R (1993). 'Variation and norm in Croatian Church Slavonic.' *Slovo* 41(13), 155–196.
- D'jačenko G (1899/1976). *Polnyj cerkovno-slavjanskij slovar'* (2 vols). Moscow: Vil'de/ Rome: JUH.
- Horbatsch O (1964). *Die vier Ausgaben der kirchenslavischen Grammatik von M. Smotryč'kyj*. Wiesbaden: Harrassowitz.
- Kraveckij A G & Pletneva A A (2001). *Istorija cerkovno-slavjanskogo jazyka v Rossii: konec XIX–XX v.* Moscow: Jazyki russkoj kul'tury.
- Pletneva A A & Kraveckij A G (2001). *Cerkovnoslavjanskij jazyk*. Moscow: Drevo dobra.
- Mathiesen R (1972). *The inflectional morphology of the Synodal Church Slavonic verb*. Columbia: Ann Arbor.
- Mathiesen R (1984). 'The Church Slavonic language question: an overview (IX–XX Centuries).' In Picchio R & Goldblatt H (eds.) *Aspects of the Slavic language question* (2 vols). New Haven: Slavica. vol. 1, 45–65.
- Miklosich F von (1862–1865/1977). *Lexicon Palaeoslovenico-Graeco-Latinum*. Vienna: Wilhelm Braumüller/ Scientia Verlag Aalen.
- Picchio R (1980). 'Church Slavonic.' In Schenker A M & Stankiewicz E (eds.) *The Slavic literary languages: formation and development*. New Haven: Slavica. 1–33.
- Plähn J (1978). *Der Gebrauch des modernen russischen Kirchenslavisch in der russischen Kirche*. Hamburg: Buske.
- Reinhart J (1990). 'Eine redaktion des kirchenslavischen Bibeltextes im Kroatien des 12. Jahrhunderts.' *Wiener Slavistisches Jahrbuch* 37, 193–241.
- Slavova T (1989). 'Preslavska redakcija na Kirilo-Metodievija starobălgarski evangelski prevod.' In Dinekov P *et al.* (eds.) *Kirilo-Metodievski studii* 6. Sofia: Bălgarska Akademija na naukite. 15–129.
- Thomson F J (1998). 'The Slavonic translation of the Old Testament.' In Krašovec J (ed.) *Interpretation of the*

Bible. Ljubljana/Sheffield: Slovenska akademija znanosti in umetnosti/Sheffield Academic Press. 605–920.
 Tolstoy NI (1988). *Istorija i struktura slavjanskix literaturnyx jazykov*. Moscow: Nauka.

Trunte N H (1998). *Slavenskij jazyk. Ein praktisches Lehrbuch des Kirchenslavischen in 30 Lektionen. Band 2: Mittel- und Neukirchenslavisch*. Munich: Otto Sagner.

Chuvash

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Chuvash (*čăvaš čělxi, čăvašla*) is the only modern representative of the Oghur (or Bulgar) branch of the Turkic language family. It is spoken in the Volga-Ural region, partly in the Chuvash Republic (*Čăvaš Respubliki*) at the ‘Great Bend’ of the Volga River. The Chuvash Republic (the capital is Cheboksary, *Šupaškar*) was established in 1990 within the Russian Federation; its forerunner was the Chuvash Autonomous Soviet Socialist Republic, created in 1925. The Chuvash have majority status in the Republic, forming nearly 70% of the population. Over three-fourths of the population regard Chuvash as their native language.

More than half of the speakers of Chuvash live outside the Republic, especially in the south and southwest parts of Tatarstan, in the central and west parts of the Bashkortostan, and in the Kuybyshev, Ulyanovsk, and Samar provinces. Speakers of Chuvash also live in other parts of Russia, in West and East Siberia, in the Far East, and in some Central Asian republics. The total number of Chuvash-speaking people is nearly 2 million. According to a law adopted in 1991, Chuvash and Russian are the official languages of the Republic. Russian is the medium of communication between nationalities and the main language of instruction. However, the efforts to maintain Chuvash are strong, even in the younger generation.

Origin and History

Parts of the old Oghur tribal confederation, originally based in the Baikal Lake region, moved west and arrived, in the mid-5th century, in the European steppe, where they established states on the Kuban, Danube, and Volga rivers. They mostly assimilated linguistically, a well-known example being the Slavization of Bulgar groups in the Balkans. At the end of the 9th century or earlier, Oghur groups settled in the Volga-Kama region, where they established the Volga Bulgar kingdom, with its center on

the middle and lower course of the Volga River. They accepted Islam as early as 922. After the destruction of this state by the Mongols in the 13th century, the Volga Bulgars and other groups of the region became subject to the Golden Horde.

Early Oghur is unknown except for the evidence found in some proper names and old loanwords. Chuvash, which was recorded for the first time in the 18th century in word lists, texts, and one grammar, is considered closely related to Volga Bulgar and other old varieties of the Oghur type. Volga Bulgar is partly known from tombstone inscriptions found on the left bank of the Volga River, dating to the 13th and 14th centuries. Several linguistic features recorded in these epitaphs do not, however, fit very well with the known features of Chuvash. It is thus still not quite clear that Chuvash is a direct descendant of Volga Bulgar. It is also unknown whether the ancestors of the Chuvash took part in the written culture of the Bulgars. There are no Volga Bulgar epitaphs on the territory of the Chuvash Republic. The fact that Chuvash is one of the very few Turkic languages that is not strongly influenced by Islam may indicate that the ancestors of the Chuvash were not affected by the Volga Bulgar Islamic culture.

Related Languages and Language Contacts

Chuvash is the result of the oldest known split within the Turkic family. Its origins reside in the language of Oghur Turkic group. Chuvash has played a key role in comparative Turkic linguistics, especially in discussions about a possible genealogical relationship of Turkic, Mongolic, and Tungusic within an Altaic language family. According to an older view, Chuvash constitutes an independent Altaic language. The hypothesis of an Altaic protolanguage relies on reconstructions on the basis of words shared by Turkic, Mongolic, Tungusic, and sometimes other languages, such as Korean and Japanese. Deviant Chuvash consonant representations have been used to reconstruct a Proto-Altaic phonology.

Chuvash words with *r* and *l* sometimes correspond to Common Turkic words with *z* and *š* (e.g., Chuvash

čul ‘stone’ vs. Common Turkic *ta:š*). This is an archaic Oghur feature. Two Samoyed words that can be traced back to **yür* ‘hundred’ and **kil* ‘winter’ have obviously been copied from Oghur words containing the same final consonants. The corresponding Chuvash words are *šer* and *xěl*, whereas other Turkic languages display forms ending in *-z* and *-š*, respectively (e.g., Turkish *yüz*, *kış*). Chuvash words with *r* and *l* sometimes have Mongolic equivalents with *r* and *l* (e.g., *čul* ‘stone’ vs. *čilayun*). Cases such as these have been used to reconstruct the special Proto-Altaic elements *r*² and *l*², which are thought to be represented by *r* and *l* in Mongolic and Chuvash, whereas they have developed into *z* and *š* in Common Turkic. Scholars who do not accept the Altaic hypothesis explain these and other correspondences by contact relationship. In this case, the assumption is that an Oghur language of the Chuvash type, with certain features, was the source of the oldest layer of Turkic loanwords in Mongolic. Tungusic, in turn, is thought to have borrowed words with these features from Mongolic.

Complex processes of linguistic assimilation have taken place in the Volga-Kama region since the 10th century. The Bulgar influence on East Finnic, Slavic, and early Kipchak Turkic was considerable. Ancestors of the Chuvash assimilated speakers of Udmurt (Votyak) and Meadow Mari (Cheremis). The assimilation of local populations led to strong substrate influences, especially from Mari. The term ‘Chuvash,’ first documented in Russian chronicles of the 16th century, originally referred to groups that also included speakers of Mari. On the other hand, the designation ‘Cheremis’ was also applied to Chuvash. After the Mongol conquest, from the 14th century on, Kipchak-speaking newcomers played an important role in the area. Speakers of Volga Bulgar were linguistically influenced by them. Parts of them assimilated Volga Bulgars and other Oghur-speaking groups, which led to substratum influence. What is known as Chuvash today remained relatively uninfluenced by the Kipchak wave. In its more recent linguistic history, however, Chuvash has been closely connected with Kipchak Turkic through massive Tatar impact.

The Written Language

Standard Chuvash is written with a Cyrillic-based alphabet that includes a few special letters. It goes back to a script system established by Ivan Jakovlev (1848–1930), which mirrors the pronunciation of the Anatri dialect. The alphabet was reformed in 1938 and has remained unchanged since then. It basically represents phonemes, and few allophones.

Distinctive Features

Chuvash shares basic linguistic features with other Turkic languages, preserving numerous so-called Common Turkic traits. It exhibits most linguistic features typical of the Turkic family (see **Turkic Languages**). It is, for example, an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. On the other hand, it strongly deviates from Common Turkic in some respects, particularly in its phonology. In the following suffix notations, capital letters indicate phonetic variation (e.g., A = ä/ě). Hyphens are used to indicate morpheme boundaries.

Phonology

Chuvash phonology displays many irregular and complicated sound changes. This is especially true of the vowels, of which correspondences with Common Turkic vowels are far from unequivocal. For instance, the Common Turkic vowel *a* is represented by *u* in words such as *ut* ‘horse’ (cf. Tatar *at*), but by *i* in words such as *pür* ‘to go’ (cf. Tatar *bar*). Chuvash possesses the reduced vowels *ă* and *ě* (e.g., *tăr* ‘to stand’, *pěr* ‘one’), which have their counterparts in neighboring languages, including Tatar, without corresponding to them in a systematic way. Originally long vowels are generally not preserved in Chuvash. In some cases, however, they are represented by diphthongs (e.g., *kěvak* ‘blue’ < *kö:k*).

Chuvash has a rather reduced consonant inventory in comparison with most other Turkic languages. Under Slavic influence, palatalized and nonpalatalized consonants are distinguished, the palatalized ones occurring before and after front vowels. Chuvash *r* sometimes corresponds to an Old Turkic interdental *ð*, as in *ura* ‘foot’ vs. *ađaq*. This is not necessarily an archaic feature. In cases such as this, early Bulgar *ð* seems to have changed into *z*, which then developed into *r* in late Bulgar.

Chuvash words are, as a rule, subject to sound harmony. The vowels of a word normally belong either to the front or to the back class. Most suffixes have a back vowel and a front vowel variant. However, some suffixes of standard Chuvash exist only in a front vowel variant: the plural suffix *-sem*, as in *ača-sem* ‘children’ (of *ača* ‘child’), and the third-person possessive suffix, as in *ivăl-ě* ‘her/his son’ (of *ivăl* ‘son’).

Grammar

The morphology exhibits certain deviations from Common Turkic patterns. There are thus exceptions from the agglutinative principles generally valid for Turkic languages (e.g., *tu* ‘mountain’ vs. *tăv-a* [mountain-DAT] ‘to the mountain’) (cf. Turkish

dağ [mountain], *dağ-a* [mountain-DAT]). Eight cases are normally distinguished for the standard language. As a result of phonetic development, the dative and accusative case markers have fused into one marker, -A. Besides the suffixless nominative, the dative-accusative, the genitive in -An, the locative in -rA, and the ablative in -rAn, Chuvash grammarians reckon with an instrumental-comitative case in -pA[IA], a privative (or abessive) case in -sAr, and a causal or purposive case in -šAn. Some scholars distinguish still more cases, e.g., a directive in -AllA. The plural suffix -sem is of unknown origin; other Turkic languages use plural suffixes of the type -lAr. The plural marker -sem follows possessive suffixes and precedes case markers (e.g., *kil-ēm-sen-čen* [house-POSS.1.SG-PL-ABL] 'from my houses'). In other Turkic languages, the plural suffix precedes the possessive markers, as in Turkish *ev-ler-im-den* [house-PL-POSS.1.SG-ABL].

The nominative forms of the personal pronouns of the first and second persons contain a proclitic deictic element *e-*, lacking in other Turkic languages: *epě* 'I', *esě* 'you (singular)', *epir* 'we', *esir* 'you (plural)' (cf. Turkish *ben, sen, biz, siz*). The reflexive pronouns of the type *xa-* plus possessive suffixes (e.g., *xam* 'I myself') are unknown in other Turkic languages. Three degrees of proximity are expressed with the demonstrative pronouns *ku* 'this', *śak(ă)* 'this there', and *śav(ă)* 'that there'. The numerals 1–10 display, besides their normal forms (*pěr(e)* 'one', *tăxxăr* 'nine'), emphatic variants with long consonants for use in isolated syntagmatic positions (*pěrre* 'one', *tăxxăr* 'nine'). Ordinals are formed with the suffix -měš, otherwise unknown in Turkic (*ikkě-měš* [two-ORD] 'second').

The Chuvash verb system does not exhibit such important deviations from the common Turkic system as has been assumed by some researchers. For example, the so-called 'aorist' (e.g., Turkish *gel-ir* [come-AOR.3.SG] 'comes, will come') is not lacking in Chuvash, but has survived as the so-called future, as in *kil-ě* [come-FUT.3.SG] 'will come'. The negated imperative is formed with a preposed particle *an* (*an pır* [NEG go.IMP] 'do not go'), whereas other Turkic languages use the negation suffix -mA with imperatives as well (e.g., Turkish *git-me* [go-NEG.IMP] 'do not go').

It has been suggested that some of these idiosyncratic Chuvash features – the deictic element *e-*, the pronouns *śakă* and *śavă*, the negative particle *an*, and the plural suffix -sem – have been copied from Mari or other Volga Finnic languages.

Lexicon

Most basic words in the Chuvash lexicon belong to the common Turkic vocabulary. Many elements have, however, been copied from other languages, mostly

from Tatar, neighboring Finnic languages, and Russian. An old layer indicates contacts with Samoyeds in southwestern Siberia. Later loans reflect the contacts with Mari in the Volga region, e.g., *pürt* 'house' < pört. Tatar dialects have exerted strong influence on the lexicon. Words of Arabic and Persian origin have mostly entered Chuvash via Tatar, but certain words were borrowed already in the Volga Bulgar period. Words of Mongolic origin have also mostly been copied from Tatar, such as *tăxta-* 'to wait' < *tuqta-* 'to stop'. There are numerous Russian loans, including *xasat* 'newspaper' < *gazeta* and *kěneke* 'book' < *kniga*. There are also many lexical elements of unknown origin.

Dialects

Modern Chuvash has two main dialects. Viryal, the 'upper' dialect, is spoken in the northern and northwestern parts of the Republic. Anatri, the 'lower' one, is spoken in the south. In the center and northeast, there is found a transitional dialect that is rather close to the lower dialect. The differences between the dialects are small. Standard Chuvash is based on Anatri dialects. Chuvash speakers living outside the Republic also speak Anatri dialects. Vowel harmony is less consistent in Standard Chuvash and Anatri than in Viryal. Tatar loans are more common in Anatri, whereas Mari and Russian loans are more common in Viryal.

Bibliography

- Andreev I A (1966). 'Čuvaškij jazyk.' In Baskakov N A (ed.) *Jazyki narodov SSSR 2. Tjurkskie jazyki*. Moskva: Nauka. 43–65.
- Ašmarin N I (1928–1950). *Thesaurus linguae tschuvashorum/Slovar' čuvaškogo jazyka*. Kazan: Čeboksary.
- Benzing J (1959). 'Das Tschuwaschische.' In Deny J *et al.* (eds.) *Philologiae turcicae fundamenta 1*. Aquis Mattiacis: Steiner. 695–751.
- Clark L (1998). 'Chuvash.' In Johanson L & Csató É Á (eds.) *The Turkic languages*. London & New York: Routledge. 434–452.
- Krueger J (1961). *Chuvash manual. Introduction, grammar, reader, and vocabulary*. Indiana University Publications, Uralic and Altaic Series 7. The Hague: Mouton.
- Róna-Tas A (ed.) (1982). *Chuvash studies*. Budapest: Akadémiai Kiadó.
- Skvortcov M I (ed.) (1982). *Čuvaško-russkij slovar'*. Moskva: Russkij Jazyk.

Relevant Website

<http://www.turkiclanguages.com> – Website with many Turkic language resources.

Classification of Languages

B J Blake, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

Classification of Languages

This article describes the principles underlying the classification of languages in this volume. Classification may be based on genetics, diffusion, lexicostatistics, or other relationships. A map (Figure 1) showing locations of major language groupings worldwide is provided.

Genetic Classification

Both professional linguists and general readers find a genetic classification the most satisfying way to group languages. This approach is one in which languages are classified into families on the basis of descent from a common ancestor. A good example is the Indo-European family of languages, which includes most of the languages of Europe, Iran, Afghanistan, and the northern part of South Asia. These languages can be shown to descend from a common ancestor, a common protolanguage. There are no records of the ancestral language, but it can be reconstructed from records of daughter languages such as Sanskrit, Ancient Greek, and Latin by using what is known as the 'comparative method'. Consider the following words for 'father': Sanskrit *pitár*, Greek *patér*, and Latin *pater*. It is possible to align the initial *ps*, the medial *ts*, and the final *rs* and reconstruct a root with the consonants *p-t-r* (the vowels require a little further examination). English is also a related language, so the word *father* should show the same consonants, but in fact the expected *p* shows as an *f* and the *t* shows up as a *th* (representing a voiced dental fricative). However, a consideration of further words shows that the *f/p* correspondence also appears in many other words, such as English *foot* against Sanskrit *pád-*, Greek *pod-*, and Latin *ped-*, and the *th/t* correspondence also shows up in other words, such as English *mother* against Sanskrit *ma:tár*, Greek *má:ter*, and Latin *mater*. We still reconstruct *p-t-k* and conclude that English has systematically changed the original stop consonants into fricatives. In fact, all the Germanic languages have done so.

Inflections as well as roots can be reconstructed. A common genitive ending in *-s* can be seen in Greek *pod-ós*, Latin *ped-is*, and English *foot's*. Proceeding in this way, we can reconstruct a good deal of the protolanguage and we can demonstrate that these languages and a score or so others are

related as members of one family, which we call Indo-European.

A language family can be represented by a tree diagram, with the branches representing subgroups. Subgroups are characterized by shared innovations, which sets them apart from other languages in the family. The Germanic branch of Indo-European (English, German, Dutch, etc.) is characterized by various consonant shifts such as *p* → *f* and *t* → *th*, as just mentioned, and by a past tense marked by a dental (or alveolar) stop, as in English *answered* or German *antwortete*. Other branches of Indo-European that can be reconstructed include Armenian, Anatolian, Celtic, Tocharian, and Italic. The Italic branch contained languages located in Italy, such as Oscan, Umbrian, and Latin. Latin was spread by conquest from Rome to a large area around the Mediterranean. It is no longer a spoken language, but it survives through its daughters, namely, French, Portuguese, Spanish, Italian, and Romanian, to mention only national languages. These languages, collectively called the Romance languages, form a sub-branch of the Italic branch of the Indo-European family. In this instance, we have records of Latin, which serves as a check on what we might reconstruct as proto-Romance. All of the Indo-European languages treated in this encyclopedia are included in the alphabetic list of families and other large groupings in the last section of this article ('Status of the Groupings Used in the Classification').

It is common in studying languages to find among them resemblances that are insufficient for the reconstruction of a protolanguage. This can be because there are insufficient data or because the languages have diverged so far that only a little evidence remains of their genetic affiliation. Where there is insufficient evidence for establishing a family or grouping families into a wider family, so that they become branches of the larger family, we can describe the languages in question as belonging to a particular stock. There can be degrees of resemblance among languages. If languages are grouped into stocks on the basis of sharing 10–20% of vocabulary, and some stocks are found to share between 5 and 10%, then these stocks can be said to belong to the one phylum.

Diffusion

In the ideal case, a number of innovations will coincide, as with the Germanic innovations mentioned previously, and a branch can be added to a tree diagram. However, all innovations, whether they are new pronunciations, new affixes, new words, or

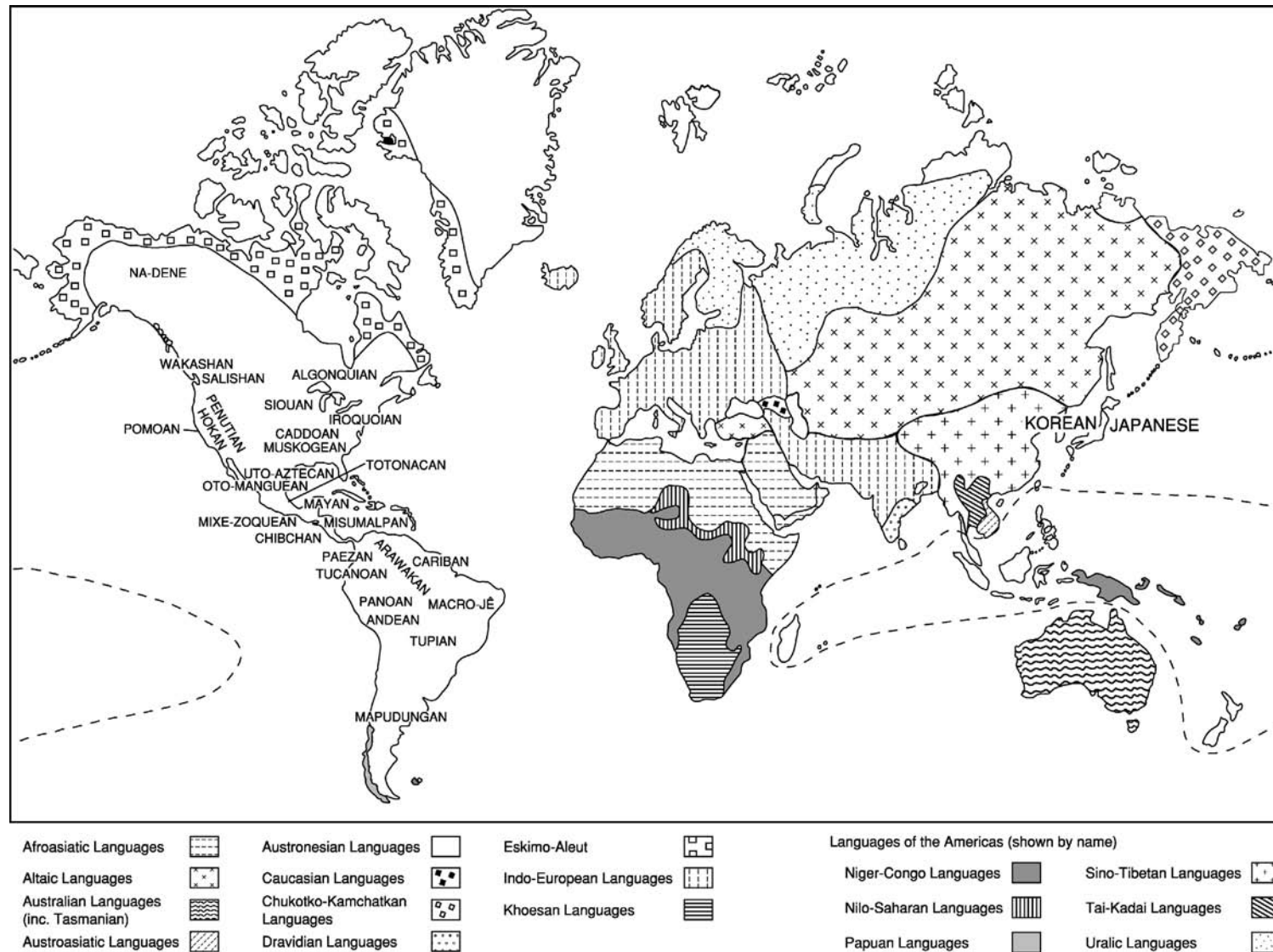


Figure 1 Locations of the major language groupings of the world, excluding the large-scale expansion of European languages such as English and Spanish over the past 500 years. The approximate locations of major concentrations are shown. In the Americas, there are many families, often with discontinuous and interlocking distributions, so the labels, indicated by name, are very approximately located.

new constructions, must start at a particular location and then spread, and different innovations can have different starting points, and the spreads can overlap. This can happen within a particular language or between languages in contact, with the result that linguists cannot always present a neat, noncontroversial tree diagram.

The diffusion of language features can be massive and widespread. Vocabulary can be borrowed from one language to another. 'Borrow' is the conventional term for the adoption of language features from another language, but no paying back is implied. Words to do with culture are most easily borrowed. English, for instance, has borrowed almost the entire learned stratum of its lexicon from French, Latin, and Greek. Similarly, Thai, Lao, and Khmer (Cambodian) have borrowed their learned stratum from Pali, a language of the Indo-Aryan branch of Indo-European. Pali is the language of Buddhism. In areas where Islam is found, languages exhibit various degrees of borrowing from Arabic. Common vocabulary is not immune from borrowing. English, for example, has borrowed *very* from French, and it has borrowed some hundreds of fairly basic words from Old Norse, including the pronominal forms *they*, *their*, and *them*. The standard tree diagram shows English as part of the West Germanic sub-branch of Germanic and Old Norse (ancestral to the modern Scandinavian languages), as representing North Germanic, but it is more realistic to think of English as a mixture, predominantly West Germanic, but with an admixture of North Germanic. And there is also the learned stratum of vocabulary already mentioned.

Though grammatical forms, particularly bound forms such as plural markers or past tense markers, are not normally borrowed, grammatical structure or patterns are relatively diffusible. It is interesting to note that most of the languages of South Asia have subject-object-verb (SOV) word order even though they belong to different language families: the Indo-Aryan branch of Indo-European, Dravidian, and the Munda branch of Austro-Asiatic. Burushaski, a language isolate spoken in northern Pakistan, is also SOV. In China, and in Laos, Thailand, and Vietnam to the south, a number of genetically diverse languages have assimilated to Chinese in having monosyllabic roots and tones. When languages converge in this way, we have a *Sprachbund* (German for 'language union'), or linguistic area. If languages were classified typologically, then various languages of different genetic provenience would be classified together because of diffusion. Vietnamese is a good example. Historically, it belongs to the Mon-Khmer branch of Austroasiatic, but it has been so

influenced by Chinese that not only has it adopted numerous Chinese words, but it has also reduced its own roots to conform to Chinese patterns and it has developed tones as in Chinese. Word order is subject-verb-object, as in Chinese.

Lexicostatistics

Linguists are not always in a position to reconstruct the relationship between languages as has been done in the case of Indo-European. Where linguists have been confronted with a number of languages that have not been studied in detail, a common situation outside Europe over the past century, they have resorted to lexicostatistics. The method is very simple. The percentages of common roots are counted using a list of 'basic' words. The theory is that basic vocabulary is resistant to borrowing, so that the percentage will give a guide to how closely languages are related. Although it is true that everyday words are less easily borrowed compared to words to do with culture (in the broadest sense), the difference is one of degree. One of the 200-word lists of basic vocabulary that has been used contains the numerals 'one', 'two', 'three', 'four', and 'five', but these can be borrowed, as in the case of the Tai languages, which have borrowed them from Chinese. The same list also contains 'animal', 'lake', and 'mountain', all of which are borrowings in English, ultimately from Latin. The problem of distinguishing roots that have been borrowed as opposed to those that have been inherited from a protolanguage is even greater when dealing with languages for which no detailed descriptions are available. Nevertheless, lexicostatistics has been widely used in the classification of the languages of various areas, including Africa, the Americas, Australia, and New Guinea. Lexicostatistics does give a good guide to the degree of similarity between languages, and on the basis of the percentages obtained it is possible to draw a hierarchical tree diagram and classify languages in terms of phylum, stock, family, branch, sub-branch, language, and dialect. However, there is no guarantee that such a tree diagram reflects the successive breaking up of protolanguages, and the terms *family*, *branch*, and *sub-branch* do not have the same meaning as these terms do when based on the comparative method.

Greenberg classified the languages of Africa and the Americas using a form of lexicostatistics. Although his classification of African languages is widely accepted and in general use, his classification of the languages of the Americas is rejected by most scholars. In this classification, all of the languages of the Americas are united in one vast Amerind family,

except for Na-Dene (mainly in northwestern part of North America) and Eskimo-Aleut in the Arctic (Greenberg, 1987).

Beyond the Language Family

As mentioned previously, there can be various degrees of resemblance between language families and the levels of relationship can be quantified lexicostatistically and described in terms of stock and phylum. But besides hypotheses of wider relationships based purely on lexicostatistics, there are hypotheses about possible relationships between families using standard techniques of reconstruction or mixtures of standard methodology and lexicostatistics. The Nostratic hypothesis is one of the boldest and most controversial approaches; largely the work of Aharon Dolgopolsky and Vladimir Illich-Svitych, the hypothesis claims that there is a macrofamily consisting of Indo-European, Semitic, Berber, Kartvelian, Uralic, Altaic, Korean, Japanese, and Dravidian (Dolgopolsky, 1998). Other work includes that of Paul Benedict, who proposed an Austro-Tai family combining Hmong-Mien (Miao-Yao), the Tai-Kadai (or Daic) family, and Austronesian. Joseph Greenberg considered that these three recognized families plus Austroasiatic form an Austric family (Ruhlen, 1991: 152–156).

Isolates

A number of languages appear to belong to no family, though in many cases they are presumably remnants of families. The following languages are examples:

- Ainu (spoken in Japan)
- Burushaski (spoken in northern Pakistan)
- Basque (spoken in the Pyrenees)
- Elamite (an extinct language of southwestern Iran; it has been claimed to be related to the Dravidian languages of southern India)
- Japanese and the Ryukyuan dialects (the latter spoken in the Ryukyu Islands of Japan)
- Ket (spoken in the Yenisei Basin, Siberia)
- Korean
- Nivkh (spoken in eastern Siberia, including Sakhalin Island)
- Sumerian (extinct language of Mesopotamia with records from the 3rd millennium B.C.)
- Yukaghir (spoken in eastern Siberia).

For most of these languages, hypotheses are put forward from time to time linking them with other languages. A number of scholars include Japanese or

Korean, or both, in the Altaic family, and some would include Yukaghir in the Uralic family.

Pidgins and Creoles

Where people find themselves in contact but without a common language, a 'pidgin' develops, which is a simplified form of language. The pidgin usually combines elements from more than one language, but in most cases the bulk of the lexicon comes from one particular language. A number of pidgins developed in the context of European colonial expansion from the 15th to the 19th centuries in places where workers, often slaves, from different language backgrounds were faced with an unfamiliar European language and in many cases unfamiliar languages of fellow workers. Where later generations learned these pidgins as their native language, the pidgins expanded to be full languages. Such languages are known as 'creoles'. In terms of classification, pidgins and creoles do not lend themselves to the hierarchical taxonomy wherein each language has a single ancestor. However, they tend to be identified in terms of which language supplies most of the vocabulary. The list of the pidgins and creoles included in this work, given in **Table 1**, shows the main source of the lexicon and where the pidgin or creole is, or was, spoken.

Table 1 Pidgins and creoles

| <i>Language</i> | <i>Main source of the lexicon</i> | <i>Location</i> |
|---------------------|-------------------------------------|--------------------------------------|
| Bislama | English | Vanuatu |
| Cape Verdean creole | Portuguese | Cape Verde |
| Fanagolo | Bantu languages (e.g., Xhosa, Zulu) | Southern Africa |
| Gullah | English | Sea Islands of South Carolina |
| Hawaiian creole | English | Hawaii, United States |
| Krio | English | Sierra Leone |
| Louisiana creole | French | Louisiana, United States |
| Mobilian jargon | Choctaw, Chickasaw | Southeastern United States (extinct) |
| Palenquero | Spanish | Colombia |
| Papamiento | Portuguese, Spanish | Aruba, Bonaire, Curaçao |
| Russenorsk | Russian, Norwegian | Arctic (extinct) |
| Sango | Ngbandi, French | Central African Republic |
| Sranan | English | Surinam |
| Tok Pisin | English | New Guinea |
| Yanito | English, Spanish | Gibraltar |

Status of the Groupings Used in the Classification

This section contains a list of language families and other groupings in alphabetic order with an indication of the status of the groupings, i.e. whether the labels represent generally accepted families, controversial families or larger entities. It should be noted that while the list covers most of the language families of the world, it is not a complete catalogue of the world's languages, which total somewhere near 5000.

Afroasiatic Languages

There are various classifications of Afroasiatic languages. The one used here recognizes six families: Ancient Egyptian and its successor, Coptic; Berber (northwest corner of Africa); Chadic (Niger and Chad); Cushitic (Somalia and eastern Sudan); Omotic (southern Ethiopia); and Semitic. Semitic has three branches. The eastern branch is represented by Akkadian, which was spoken in Mesopotamia from the 3rd to the 1st millennium B.C. The southern branch is represented by the Ethiopian languages (Amharic, Tigrinya, and the extinct Ge'ez). The central branch, which is centered around the eastern end of the Mediterranean, includes the dead languages Phoenician, Syriac, and Ugaritic, plus Aramaic, a language in which parts of the Bible are written and which is still spoken; Hebrew, which has been brought back to life as the language of Israel; and Arabic, which, as the language of Islam, has spread over northern Africa and the Middle East.

Altaic Languages

Altaic is a widely, though not universally, accepted language family covering three branches: Turkic, Mongolic, and Tungusic, represented in this work by Evenki. The Turkic languages, which include Turkish, extend across from the Balkans through Turkey across central Asia to Siberia. The Mongolic languages are centered on Mongolia and the Tungusic languages in Siberia and northern China. If Altaic is rejected as a family, then we have three separate families rather than three branches of a family. These languages are typologically similar in that they are agglutinative, and they represent the classic SOV word-order type with SOV word order, postpositions, and preposed genitives. Some linguists would include Japanese and/or Korean in the Altaic family.

Australian Languages

The languages of the Australian mainland look as if they are related, but no detailed reconstruction of a

protolanguage has been undertaken and it is unlikely that such a reconstruction will be possible. These languages have been classified lexicostatistically, i.e., by counting percentages of common vocabulary. This classification currently recognizes about a score of lexicostatistical families, with one of them, Pama-Nyungan, covering most of the mainland. Some genetic groupings are recognizable within Pama-Nyungan, and some of the other lexicostatistical families can be shown to be true families, such as the Tangkic family, which includes Kayardild, and West Barkly, which includes Wambaya. Tiwi is the sole member of the Tiwian family. Dixon (2002: 674) suggests that the similar-looking Daly group of languages (represented in this work by Ngan'gityemerri) is an areal group rather than a genetic one. Records of the extinct Tasmanian languages consist almost entirely of amateur word lists. These show very few resemblances to the languages of the mainland. Joseph Greenberg classified the Tasmanian languages, the Papuan languages, and the languages of the Andaman Islands in an Indo-Pacific phylum (Ruhlen, 1991). This grouping has been disregarded by almost all other linguists.

Austroasiatic Languages

The Austroasiatic classification comprises two branches: the Munda languages of northeast India, which includes Santali, and the more scattered Mon-Khmer branch, which includes Mon (southeastern Myanmar (Burma)), Khmer (or Cambodian, the official language of Cambodia), Khasi (northeast India), Wa (southwest Yunnan, China), and Vietnamese. Vietnamese is interesting from the point of view of classification. It has been so influenced by Chinese that as well as borrowing large numbers of Chinese words, it has reduced the form of roots and developed tones so that the language looks like a Chinese language.

Austronesian Languages

The Austronesian language family contains over 1000 languages. In the most widely used classification, there are four branches, Paiwanic, Tsouic, Aytalic, and Malayo-Polynesian. The first three are the indigenous languages of Taiwan and are collectively known as the Formosan languages. The extra-Formosan languages, which are assumed to have emanated from Taiwan, make up the Malayo-Polynesian branch, which is spread from Madagascar in the western Indian Ocean, where Malagasy is spoken, to Easter Island in the eastern Pacific. Oversimplifying somewhat, we can consider there are three sub-branches: western, which takes in the languages of the Philippines, Indonesia, and Malaysia as well as

Malagasy and Hawaiian; central, represented in this work by the Flores languages and Malukan languages; and Oceanic, which covers languages such as Fijian, Maori, Samoan, Tahitian, and Tongan.

Caucasian Languages

The languages of the Caucasus comprise a South Caucasian or Kartvelian family, represented here by Georgian, and the North Caucasian languages, with a northwestern sub-branch, represented here by Abkhaz, and a northeastern sub-branch, represented by Lak. It is not quite certain that the northwestern branch and northeastern Branch are branches of a single family, and it is even more uncertain that South Caucasian and North Caucasian families form a genetic group, but the label 'Caucasian languages' is useful since the two groups share some features and are all quite distinct from surrounding languages.

Chukotko-Kamchatkan Languages

This is a small family of languages spoken on the Chukotka and Kamchatka peninsulas of Siberia.

Dravidian Languages

This language family is concentrated in southern India. Some branches are recognizable. Dravidian proper includes Gondi, Kurukh, and Telegu; the southern branch includes Kannada, Malayalam, Tamil, and Toda, and the northwestern branch includes Brahui.

Eskimo-Aleut

The Eskimo-Aleut language family has two primary branches. The Aleut branch is spoken in the Aleutian Islands and the Eskimo languages are found in Siberia, Alaska, Canada, and Greenland. The latter branch is represented here by Inupiaq and West Greenlandic.

Indo-European

Indo-European is the most widely studied of all language families and has a well-articulated sub-grouping based on the comparative method, though details of the classification are subject to dispute from time to time. This family of languages contains a number of branches containing a single language (or group of dialects), namely, Albanian, Armenian, Hellenic (Greek), and two dead languages, records of which came to light only in the 20th century. One dead language is Hittite, which was spoken in Anatolia (modern Turkey). There are records of Hittite from the latter part of the second millennium B.C. The other dead language, Tocharian, the

easternmost Indo-European language, was spoken in what is now the Xinjiang province of western China. There are records of Tocharian from the period 500–700 A.D.

Among other branches are the following Indo-European languages:

- Baltic contains Lithuanian and Latvian, and Slavic, the earliest records of which are in Old Church Slavonic and date from the 11th and 12th centuries. Modern Slavic languages include Polish, Sorbian, Czech, and Slovak (western sub-branch); Bulgarian, Macedonian, Slovene, and the 'Serbian-Croatian-Bosnian complex' (southern sub-branch); and Russian, Belorussian, and Ukrainian (eastern sub-branch). Some linguists would classify Baltic and Slavic as sub-branches of a Balto-Slavic branch.
- Celtic is usually divided into two sub-branches: the Brythonic branch, which contains Breton, Cornish, Welsh, and possibly Pictish, about which little is known, and the Goidelic branch, which contains Scots Gaelic.
- Germanic contains three sub-branches. The eastern sub-branch is represented by the extinct Gothic; the northern sub-branch, by the Scandinavian languages (Danish, Icelandic, Norwegian, Swedish); and the western sub-branch, by German (including High German, Yiddish, and Low German), Frisian, Dutch, and its South African derivative, Afrikaans, and various forms of English, including Scots.
- Indo-Iranian is a large branch containing two large sub-branches, Indo-Aryan (or Indic) and Iranian. Indo-Aryan covers Sanskrit, the language of the Hindu sacred texts; Pali, the language of the Hinayana Buddhist canon; plus Bengalic, the Dardic languages, Dhivehi, Domari, Gujarati, Hindi, Hindustani, Kashmiri, Lahnda, Marathi, Nepali, Punjabi, Sindhi, Sinhala, and Urdu, all of which are spoken in India, Pakistan, and Bangladesh, plus Romani, the language of scattered Gypsy communities. Iranian covers Avestan, the language of the Zoroastrian scriptures, plus Bactrian, Baluchi, Chorasmian, Khotanese, Kurdish, Ossetic, Pahlavi (Middle Persian), Pashto, Persian, Sogdian, and Tajik.
- Italic contains a number of extinct languages of Italy, one of which, Latin, was spread via the political dominance of Rome. The descendants of Latin, known collectively as the Romance Languages, include several national languages (French, Italian, Portuguese, Romanian, and Spanish) as well as Catalan (northeastern Spain), Galician (northwestern Spain), Jerriais (Jersey), Occitan (southern France), Rhaeto-Romance (eastern Switzerland and northeastern Italy), and Sardinian.

Khoesaaan Languages

The Khoesaaan group of languages is spoken by the Khoekhoe and San peoples of southern Africa. The group is often described as having three branches, but the branches are probably separate families. Two languages of northern Tanzania, Hadza and Sandawe, are also included in the group in most reference works, but it is not clear that they are genetically related to any of the southern families.

Languages of the Americas

As mentioned in the preceding section on lexico-statistics, Joseph Greenberg classified all of the languages of the Americas in one vast Amerind family, except for Na-Dene (mainly in northwestern part of North America) and Eskimo-Aleut in the Arctic. This classification is generally rejected and most scholars would recognize some scores of separate families in Greenberg's Amerind, though allowing that some of these can be grouped into stocks. We have followed a widespread convention of breaking up the languages of the Americas into three geographical regions: North America, Central America, and South America. This is largely to reduce a very large area to manageable chunks. We have considered Eskimo-Aleut separately from the languages of the Americas since it is not confined to North America.

Languages of North America

- The Algonquian languages are found in the eastern part of North America and westward into Alberta and Montana, and the Ritwan languages (Wiyot and Yurok) are found in northern California. Mithun (1999: 327) recognizes Eastern Algonquian, Central and Plains Algonquian, and Ritwan as branches of an Algic family. Algonquian is represented in this work by Cree and Michif. Michif is a creole, but, unlike most creoles, it did not arise from a pidgin. It retains the complex verbal morphology of Cree, and noun phrases show distinctions of number, gender, and definiteness, as in French.
- The Caddoan language family belongs to the Great Plains of the midwestern United States.
- The Hokan group of languages is centered in California. It is not established that these languages form a family. Among the Hokan languages is the Pomoan family of northern California.
- The Iroquoian language family of southeastern Canada and the eastern United States is represented in this work by Oneida (Northern Iroquoian) and Cherokee (Southern Iroquoian).

- The Keres language consists of a number of dialects spoken in New Mexico.
- The Muskogean language family of the southeastern United States includes Choctaw (Mississippi) and Creek (Alabama and Georgia).
- The Na-Dene language family includes Tlingit, Eyak, and the large Athapaskan branch. Most of these languages belong to Alaska and western Canada, but there is an enclave of Athapaskan in the southwest of the United States. Navajo (Navaho) is spoken in Arizona, New Mexico, and Utah.
- The Penutian group of languages or stock belongs to the west of North America, from British Columbia to California.
- Languages of the Salishan family are spoken in British Columbia and the northwest of the United States.
- The Siouan family of languages covered a vast area of the Great Plains and included Crow, Lakota, and Omaha-Ponca.
- The Wakashan language family is mainly from Vancouver Island, British Columbia, and is represented by Nuuchahnulth (Nootka).

Languages of Central America

- Languages of the Chibchan family are spoken in Nicaragua, Costa Rica, Panama, western Colombia, and Ecuador, and the Paezan languages are spoken in Colombia.
- The Mayan family of languages is spoken in southeastern Mexico and Guatemala.
- The Misumalpan language family is found in western Honduras and western Nicaragua.
- The Mixe-Zoquean language family is found in southern Mexico.
- The Oto-Manguean language, represented here by Zapotecan, is found in Southern Mexico.
- The Uto-Aztecan language family is found mainly in the southwest of the United States and Mexico, but extends as far north as Idaho. This family includes Cupeño, Hopi, Tohono O'odham, and Nahuatl, the language of the Aztec civilization.

Languages of South America

- The most widely spoken native language of South America is Quechua. It is spoken in Peru, Ecuador, and Bolivia, extending north into Colombia and extending south into northern Chile and northwestern Argentina. It shares similarities with Aymará and the two are sometimes grouped in an Andean family, but this is not generally accepted, since it is not agreed whether the resemblances are genetic or arise from contact.

- The large Arawak language family is widespread, ranging from Honduras in Central America to Brazil in South America, and formerly to Paraguay and Argentina. The Arawak language in this work is Tariana, of Brazil.
- The large Carib language family is found in Brazil and the countries of South America north of Brazil.
- The Choco language family is found in Brazil.
- The languages of the Panoan family are found in Peru and neighboring parts of Bolivia and Brazil.
- Macro-Jê is a grouping of languages that have been considered to be related to the Jê family. These languages are located in Brazil.
- The Mapudungan language is spoken in Chile and Argentina. It has no clear genetic affiliation.
- The Tucanoan language family is found in western Brazil and neighboring parts of Colombia, Ecuador, and Brazil.
- The Tupian language family is located in Brazil. The Tupí-Guaraní sub-group is also found in Brazil, but various members of the sub-group are found in Bolivia, Paraguay, and Argentina. Guaraní is an official language of Paraguay, along with Spanish.

Niger-Congo Languages

This is a very large language family, with about 1000 members. It is spread over southern Africa. There are various classifications, including some that are hierarchical with several levels. We have adopted a flat classification with eight branches:

- The Kordofanian group of languages is spoken in Sudan. In some classifications, a Niger-Kordofanian family is recognized, with Kordofanian and Niger-Congo as the primary branches.
- The Atlantic Congo language sub-group is located in the far west of Africa from Liberia to Senegal. It includes Fula and Wolof.
- Languages of the Kru sub-group are spoken in Ivory Coast and Liberia.
- The Mande language sub-group is found from Senegal to Burkina Faso (Upper Volta) and Ivory Coast.
- The Gur (Voltaic) language sub-group is spoken in Mali, Burkina Faso, and Ghana, and extends east into Nigeria. In some classifications, Dogon is not assigned to any branch; in others, it is assigned to the Gur sub-branch.
- The Kwa sub-group of languages extends from Liberia to Nigeria.
- The Benue-Congo language sub-group covers a very large part of southern Africa. This branch includes Efik, Yukuben, and Mambila. The very large Bantu language group, which includes

Kikuyu, Kinyarwanda, Nyanja, Shona, Swahili, Xhosa, and Zulu, is a sub-group of Benue-Congo and hence a sub-sub-group of Niger-Congo.

- The Adama-Ubangi language sub-group is spoken in a band running across Africa from Nigeria to Sudan.

Nilo-Saharan

The languages of the Nilo-Saharan family are found mainly in northeastern and north-central Africa. They include Dinka, Kanuri, Luo, and the Songhay languages.

Papuan Languages

The label 'Papuan' has no genetic significance. It is defined negatively as the non-Austronesian languages of New Guinea and surrounding islands. It covers about 750 languages in New Guinea and another 50 or so on neighboring islands from Timor to the Solomons. These languages can be classified into 23 families and 10 isolates. One very large family, the Trans-New Guinea family, covers most of New Guinea and is also found on some of the neighboring islands. It contains a number of branches, including the Madang languages. Other families include Sepik, represented in this work by Manambu of the Ndu subgroup, Skou, Torricelli, and West Papuan. Also included in this work is an article on several of the Papuan languages of the central Solomons.

Sino-Tibetan

The Sino-Tibetan languages include the Sinitic family and Tibeto-Burman. Sinitic can be equated with Chinese, but Chinese is popularly understood to be a single language, whereas in fact it is more like a family of languages, one of which, Mandarin Chinese, is the standard, based largely on the Beijing dialect. Tibeto-Burman takes in a number of genetically related languages, including Tibetan and Burmese, but there is no consensus about the details of the classification. Whether Tibeto-Burman and Sinitic are genetically related is not agreed, but there are some apparent cognates.

Tai Languages

The Tai, or Daic, language family is centered in Laos and Thailand and includes the national languages of these two countries, Lao (or Laotian) and Thai. The family is also represented in Burma, southern China, northern Vietnam, and on Hainan Island in the Gulf of Tonkin. Lao and Thai are mutually comprehensible. A purely linguistic classification would recognize a chain of Tai dialects across the two countries that included the national languages.

Uralic Languages

The Uralic languages are a family of languages spoken in northeastern Europe, extending across northern Russia into northwestern Siberia. There are two major branches, the Samoyed branch, represented in this work by Nenets, spoken in northern Russia, and Finno-Ugric, which includes Estonian, Finnish, and Saami (spoken in northern Norway, Sweden, and Finland), as well as Hungarian, the national language of Hungary, which is separated from the rest of the family. Some would include Yuhaghir in the Uralic family, others would combine Uralic and Altaic into a larger family.

Language Classification**Afroasiatic Languages**

- Ancient Egyptian and Coptic
- Berber Languages
- Chadic Languages
 - Hausa
- Cushitic Languages
 - Highland East Cushitic Languages
 - Oromo
 - Somali
- Omotiic Languages
 - Wolaitta
- Semitic Languages
 - Eblaite
 - Eastern
 - Akkadian
 - Central
 - Arabic
 - Aramaic
 - Hebrew, Biblical and Jewish
 - Hebrew, Israeli
 - Jewish languages
 - Maltese
 - Phoenician and Punic
 - Syriac
 - Ugaritic
 - Southern
 - Ethiopian Semitic Languages
 - Amharic
 - Ge'ez
 - Tigrinya

Altaic Languages

- Mongolic Languages
- Tungusic Languages
 - Evenki
- Turkic Languages
 - Azerbaijani
 - Bashkir
 - Chuvash
 - Kazakh
 - Kirghiz
 - Tatar
 - Turkish
 - Turkmen

Language Classification (cont.)

- Uyghur
- Uzbek
- Yakut

Australian Languages

- Pama-Nyungan
 - Arrernte
 - Gamilaraay
 - Guugu Yimidhirr
 - Jiwarli
 - Kalkutungu
 - Kaytety
 - Morrobalama
 - Pitjantjatjara/Yankunytjatjara
 - Warlpiri
- Daly
 - Ngan'gi
- Tangkic
 - Kayardild
- Tiwian
 - Tiwi
- West Barkly
 - Wambaya

Austroasiatic Languages

- Mon-Khmer Languages
 - Northern
 - Khasi
 - Vietnamese
 - Wa
 - Eastern
 - Khmer
 - Southern
 - Mon
- Munda Languages
 - Santali

Austronesian Languages

- Formosan Languages
- Malayo-Polynesian Languages
 - Western
 - Balinese
 - Bikol
 - Cebuano
 - Hawaiian
 - Hiligaynon
 - Ilocano
 - Javanese
 - Kapampangan
 - Madurese
 - Malagasy
 - Malay (Malaysian and Indonesian)
 - Niuean
 - North Philippine Languages
 - Riau Indonesian
 - Samar-Leyte
 - South-Philippine Languages
 - Tagalog
 - Central
 - Flores Languages
 - Malukan Languages
 - Oceanic
 - Fijian

Language Classification (cont.)

Maori
Tahitian
Tamambo
Vures

Caucasian Languages

Abkhaz
Georgian
Lak

Chukotko-Kamchatkan Languages**Dravidian Languages**

Brahui
Kannada
Kurukh
Malayalam
Tamil
Telugu
Toda

Hmong-Mien Languages**Indo-European Languages**

Albanian
Anatolian Languages
 Hittite
Armenian
Balto-Slavic Languages
Baltic Languages
 Latvian
 Lithuanian
Slavic Languages
 Belorussian
 Bulgarian
 Church Slavonic
 Czech
 Macedonian
 Old Church Slavonic
 Polish
 Russian
 'Serbian-Croatian-Bosnian Linguistic Complex'
 Slovak
 Slovene
 Sorbian
 Ukrainian
Celtic Languages
 Breton
 Cornish
 Irish
 Pictish
 Scots Gaelic
 Welsh
Germanic Languages
 Afrikaans
 Danish
 Dutch
 English, Early Modern
 English: African American Vernacular
 English: Middle English
 English, Later Modern
 English in the present day
 English: Old English
 English, World

Language Classification (cont.)

German
Gothic
Luxembourgish
Norse and Icelandic
Norwegian
Old Icelandic
Scots
Swedish
Yiddish
Hellenic
 Greek, Ancient
 Greek, Modern
Indo-Iranian Languages
 Indo Aryan Languages
 Assamese
 Bengali
 Dardic
 Kashmiri
 Dhivehi
 Domari
 Gujarati
 Hindi
 Hindustani
 Lahnda
 Marathi
 Nepali
 Nuristani Languages
 Pali
 Punjabi
 Romani
 Sanskrit
 Sindhi
 Sinhala
 Urdu
 Iranian Languages
 Avestan
 Bactrian
 Balochi
 Chorasmian
 Khotanese
 Kurdish
 Ossetic
 Pahlavi
 Pashto
 Persian, Modern
 Persian, Old
 Sogdian
 Tajik Persian
 Italic Languages
 Latin
 Romance Languages
 Catalan
 Franglais
 French
 Galician
 Italian
 Jerriais
 Occitan
 Portuguese
 Rhaeto Romance
 Romanian
 Spanish
 Tocharian

Language Classification (cont.)**Khoesaaan Languages**

Khoesaaan Languages

Languages of the Americas**Languages of North America**

Algonquian and Ritwan Languages

Cree

Mitchif

Caddoan Languages

Eskimo-Aleut

Inupiaq

West Greenlandic

Hokan Languages

Pomoan Languages

Iroquoian Languages

Oneida

Keres

Muskogean Languages

Choctaw

Creek

Na-Dene Languages

Navaho

Penutian Languages

Salishan Languages

Siouan Languages

Crow

Lakota

Omaha-Ponca

Wakashan Languages

Nuuchahnulth

Languages of Middle America

Chibchan

Mayan Languages

Misumalpan

Mize-Zoquean Languages

Oto-Manguean Languages

Zapotecan

Totonacan Languages

Uto-Aztecan Languages

Cupeño

Hopi

Nahuatl

Tohono O'odham

Languages of South America

Andean Languages

Aymará

Quechua

Arawak Languages

Tariana

Cariban Languages

Choco Languages

Chibchan (see Languages of Middle America)

Macro-Jê Languages

Mapudungan

Panoan

Tucanoan Languages

Tupian Languages

Guarani

Language Classification (cont.)**Niger-Congo Languages**

Kordofanian Languages

Mande Languages

Atlantic Congo Languages

Fulfulde

Ijo

Wolof

Dogon

Gur Languages

Kru Languages

Adamawa-Ubangi

Kwa Languages

Akan

Ewe

Yoruba

Benue-Congo Languages

Efik

Mambila

Bantu Languages

Gikuyu

Kinyarwanda

Luganda

Nyanja

Shona

Swahili

Xhosa

Zulu

Southern Bantu Languages

Nilo-Saharan Languages

Dinka

Kanuri

Luo

Songhay Languages

Papuan Languages

Central Solomon Languages

Sepik Languages

Manambu

Skou Languages

Torricelli Languages

Trans New Guinea Languages

Madang Languages

West Papuan Languages

Pidgins and Creoles

Bislama

Cape Verdean Creole

Fanagolo

Gullah

Hawaiian Creole English

Hiri Motu

Krio

Louisiana Creole

Mobilian Jargon

Palenquero

Papamientu

Russenorsk

Sango

Tok Pisin

Tsotsi Taal

Yanito

Language Classification (cont.)**Sino-Tibetan Languages**

Sinitic Languages
Chinese
Tibeto-Burman Languages
Burmese
Karen Languages
Tibetan

Tai Languages

Lao
Thai

Uralic Languages

Estonian
Finnish
Hungarian
Nenets
Saami

Language isolates and Languages of disputed affiliation

Ainu
Basque
Burushaski
Elamite
Etruscan
Hurrian
Japanese
Ryukyuan
Ket
Korean
Nivkh
Sumerian
Yukaghir

Artificial Languages

Esperanto

Bibliography

- Comrie B (ed.) (1987). *The world's major languages*. London: Croom Helm.
- Comrie B (1981). *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Dixon R M W (2002). *The languages of Australia*. Cambridge: Cambridge University Press.
- Dixon R M W & Aikhenvald A Y (eds.) (1999). *The Amazonian languages*. Cambridge: Cambridge University Press.
- Dolgopolsky A (1998). *The Nostratic macrofamily and linguistic paleontology*. Cambridge: McDonald Institute for Archaeological Research.
- Greenberg J H (1987). *Language in the Americas*. Stanford: Stanford University Press.
- Grimes B F (2000). *Ethnologue: languages of the world* (14th edn.). Dallas: Summer Institute of Linguistics (<http://www.ethnologue.com>).
- Mithun M (1999). *The languages of native North America*. Cambridge: Cambridge University Press.
- Ruhlen M (1991). *A guide to the world's languages. Volume 1: classification* (rev. edn.). Stanford: Stanford University Press.
- Suárez J A (1983). *The Mesoamerican Indian languages*. Cambridge: Cambridge University Press.

Cornish

P J Payton, University of Exeter, Cornwall, UK

© 2006 Elsevier Ltd. All rights reserved.

Cornish is a member of the Brythonic branch of the Celtic languages; it is related closely to Welsh and, especially, Breton and less closely to the Goidelic group comprising Irish, Scots Gaelic, and Manx. It emerged as a recognizably distinct language in the early medieval period. Although Anglo-Saxon traders or settlers had brought English into the far north-eastern tip of Cornwall, place-name evidence shows that, by the year 1200, Cornish was still spoken over the greater part of Cornwall. By 1500, the language had retreated westward to the River Fowey/River Camel line in mid-Cornwall; it was then spoken in a little more than half of the territory by approximately

30 000 people. Thereafter, decline was swift, with a core of about 5000 speakers left by 1700 and only a handful by the mid-18th century. Dolly Pentreath, popularly supposed to be the last Cornish speaker, died in 1777, but she was certainly survived by others. John Davey, who died in 1891, was said to have been able to converse in Cornish on a few simple matters, and counting rituals in Cornish survived in fishing communities until the 1920s and 1930s.

Cornish is divided into three periods: Old Cornish (from the 9th to the 13th centuries), Middle Cornish (from the 13th to the mid-16th centuries) and Late or Modern Cornish for the final period. The main corpus of literature survives from the Middle Cornish period, notably the *Ordinalia*, *Beunans Meriasek* ('The Life of St. Meriasek'), *Greans an Bys* ('The Creation of the World'), and the recently

discovered (2002) *Beunans Ke* ('The Life of St Kea'). The Reformation ensured that these plays were seen as 'subversive,' and the Cornish rebellions of 1497 and 1549 – the latter explicitly against the introduction of English in Cornish church services – meant that the Cornish language too had become subversive. The post-Reformation loss of contacts with Brittany deprived Cornish of an important cultural resource, including access to a mutually intelligible language (Breton).

Although a group of Cornish scholars did what they could to encourage the survival of the language in the late 17th and early 18th centuries, when the Celtic scholar Edward Lhuyd visited Cornwall around 1700 he found that Cornish was spoken only in 25 parishes in the far west. Antiquarian interest continued throughout the 18th and 19th centuries, with edited versions of the several of the plays published, but it was not until the publication of Edward Jenner's *Handbook of the Cornish language* in 1904

that a serious revivalist movement emerged. In the inter-war period, Robert Morton Nance produced a synthesis that he dubbed 'Unified Cornish.' This remained the standard for Cornish learners until the late 1980s when competing forms based on Late/Modern Cornish and a phonemic form of Middle Cornish emerged. However, despite this dissension, Cornish was one of the indigenous British languages recognized by the British government in 2003 under the terms of the Council of Europe's Charter on Minority Languages.

Bibliography

- Payton P (2000). 'Cornish.' In Glanville P (ed.) *Languages in Britain and Ireland*. Oxford: Blackwell. 109–119.
- Spriggs M (2003). 'Where Cornish was spoken and when.' In Payton P (ed.) *Cornish studies: Eleven*. Exeter: University of Exeter Press. 228–269.

Cree

H C Wolfart, University of Manitoba, Winnipeg, Manitoba, Canada

© 2006 Elsevier Ltd. All rights reserved.

The Cree language exhibits an extraordinarily rich morphology, traditionally compared in its profusion to that of Ancient Greek.

Inflexion

The epistemological import of a statement, for example, may not only be indicated by particles such as *êsa* 'reportedly' or *iska* 'by dream or revelation' or through direct quotation, but also inflexionally, e.g., in the dubitative form of the changed conjunct verb *wêhtinâkwê* 'there he must have obtained him' in such sentences as ... *tânitê mîna wêhtinâkwê askihkwa*. '... I wonder where he got a pail.' The stem *ohtin-* 'thus or there obtain s.o.', which requires an antecedent (*tânitê* 'whence'), is followed by the thematic suffix *-â-*, specifying a proximate (central) agent and an obviative (noncentral) patient of animate gender. The dubitative suffix *-kwê*, finally, combines with the ablaut (apophony) affecting the initial vowel of the word to express subordination to the interrogative and evidential modality.

Cree verbs are inflected in four major paradigms, with the stems themselves typically grouped into two

derivational pairs. Stative verbs differ by the gender of the agent, *mihkwâ-* 'be red (inanimate),' *mihkosi-* 'be red (animate),' while transitive verbs are distinguished by the gender of the patient: *pakamah-* 'strike s.t. (inanimate),' *pakamahw-* 'strike s.o. (animate).' Verbs of this last class specify both agent and patient inflexionally.

Their overall complexity aside, the inflexional paradigms of Cree are also subject to substantial dialect variation, both in particular endings and in entire paradigmatic dimensions.

Number in Inflexion and Derivation

In the expression of grammatical categories, inflexion and derivation complement each other. The basic distinction of number, for example, is that of singular and plural expressed inflexionally, e.g., *iskwêw* 'woman; a woman; one woman' vs. *iskwêwak* 'women (more than one).' The singular is unmarked and in elevated prose may be used collectively, as in *kôsisiminaw* 'our grandchildren [lit. 'our grandchild']'; in a sentence like *kîspin êkâ iskwêw ôta kîpakitînikowisit* 'if women [lit. 'Woman'] had not been put here [on earth] by divine powers,' both the noun *iskwêw* and the verb (in the simple [unchanged] conjunct mode and with the third-person suffix *-t-*) show the singular. Quantifiers such as *mihcêt* 'many'

heighten the literary effect of this device, e.g., *mihcêṭ namôy kiskêyihṭam* ‘many do not [lit. ‘does not’] know this.’ Reciprocal stems construed as singulars, e.g., *ayisiyiniw k-âyimôhtot* ‘that people [lit. ‘a person’] should gossip about one another,’ are a mark of high rhetoric.

The number system of Cree is remarkable for its range of associative plural constructions; for example, a first person plural verb accompanied by a singular noun (here flanked by the demonstrative *awa*) indicates a conjoint noun phrase including the first person: *kâ-sipwêhtêyâhk awa nisîmis awa* ‘when we took off, my little sister [and I].’ A third person plural verb construed with a singular noun is interpreted as including the extended family hunting band of the person specified by the noun: *ita nôhtâwiy k-âyâcik* ‘where my father [and his people] live’; this construction is rare among the languages of the world.

The opposition of singular and plural is neutralized for both nouns and verbs in the obviative (noncentral) third person, e.g., *sîsîpa tahkonêw* ‘she carried a duck/ducks,’ where the number of the patient is not specified in the verb form and the noun *sîsîpa* itself is number-indifferent. In such contexts, it is not uncommon for the verb stem to be reduplicated, e.g., *ê-kî-miyikoyâhk mâna sîsîpa ka-tâh-tahkonâyâhkik*. ‘he used to give us [each several] ducks to carry.’ With the long vowel *-â-* and strong devoicing (*-h-*) before what is treated as a word-boundary (indicated by a word-internal hyphen in the standard roman orthography), this highly productive type of ‘heavy’ reduplication expresses iteration or distributive action, which here takes the place of inflexionally marked plurality.

High literary style also exploits a much less productive type of disyllabic reduplication, which does not introduce a word boundary; the reduplication syllable is ‘light’ (with the short vowel *-a-*), and the initial syllable of the stem itself shows lengthening of the vowel in accordance with rules that are largely those of the ablaut pattern illustrated in our opening example. This more archaic type of reduplication typically appears with paired referents, e.g., *ê-mamêhkwâpicik* ‘they had their eyes painted red’ (cf. *mihkwâpi-* ‘have a red eye’), suggesting a kind of dual marking not otherwise reported for the Algonquian languages.

Derivation

The derivational morphology of Cree easily matches the exuberance of the inflexional paradigms. In the formation of primary stems from roots (or initials), optional medials and obligatory finals, the stems

corresponding to the four major verb paradigms (with their pairings by gender) are thrown into sharp relief:

pakâsim- ‘boil s.o. (e.g., a rabbit, a beaver) in water’
pakâhtâ- ‘boil (it) (e.g., bones, clothes) in water’
pakâso- ‘be boiled in water (animate; e.g., a rabbit, a beaver)’
pakâhtê- ‘be boiled in water (inanimate; e.g., bones, clothes)’

(The root on which these stems are based refers to immersion; cf. the secondary stem *pakâsimo-* ‘immerse oneself in water; swim.’) Semantic sets are not necessarily fourfold or symmetrical, nor are they confined to a canonical stem, occasionally showing suppletion instead, e.g., *mow-* ‘eat s.o. (a duck, bread),’ *mîci-* ‘eat (it).’

The initial constituent of the stem defines one paradigmatic set; e.g., *âyimôt-* ‘speak unguardedly about s.t., gossip about s.t.,’ *âyimôm-* ‘speak unguardedly about s.o., gossip about s.o.’ (and secondary stems *âyimômiso-*, *âyimôhto-* ‘speak unguardedly, gossip about oneself, one another’), *âyimisi-* ‘be of difficult disposition,’ *âyimî-* ‘have a difficult life,’ *âyiman-* ‘be difficult.’ An equally prominent set is defined by the final, e.g., the *-ôt-* which recurs in *mâmiskôt-* ‘expound s.t.,’ *tâhkôt-* ‘discourse upon s.t.,’ *tipôt-* ‘discuss s.t. with authority.’ In less specialized domains, initials such as *pîmi-* ‘in linear progression,’ *sîpwê-* ‘departing,’ *tako-* ‘arriving’ and finals such as *-ohtê-* ‘walk,’ *-pahtâ-* ‘run’ cooccur freely.

In secondary derivation, full stems give rise to further stems, e.g., the transitive *pakamahw-* ‘strike s.o.’ to the reciprocal *pakamahoto-* ‘strike one another’ or the inagentive verb of suffering *pakamahokowisi-* ‘be struck by divine force’; or the parallel stem *pakamah-* ‘strike s.t.’ to the noun stem *pakamahikanis-* ‘club, hammer.’ Noun stems such as *maskisin-* ‘moccasin, shoe’ or *nîmihitowin-* ‘dance, dance ceremony’ (itself derived from the reciprocal verb stem *nîmihito-*) yield such verb stems as *maskisinhkê-* ‘make moccasins, make shoes’ or *nîmihitowinhkê-* ‘hold a dance ceremony; give a dance.’ Some highly transitive verbs permit the formation of patient nouns, e.g., *misw-* ‘shoot s.o., wound s.o.,’ *miswâkan-* ‘wounded person,’ which in turn is the base for *miswâkaniwi-* ‘be wounded.’

Recursive suffixation is complemented by a highly productive pattern of deriving noninitial nominals from stems. While some of these deverbative medials closely resemble the full noun, others are distinct; the medials *-ihkomân-*, *-astimw-*, *-askisin-*, for instance, vary more or less obviously from the stems *môhkomân-* ‘knife,’ *atimw-* ‘dog; horse’ or *maskisin-* ‘moccasin, shoe,’ as in *mistihkomân* ‘big knife’

or *manihkomânê*- ‘take a knife,’ *mihcêtwastimwê*- ‘have many horses’ or *kikaskisinê*- ‘wear moccasins, wear shoes.’

Noun Incorporation

Transitive stems with animate patients may function as the base of an overtly incorporative stem consisting, for instance, of the stem *ohtin*- ‘thus or there obtain s.o.’ as the initial constituent, the derived medial *-iskwêw*- ‘woman’ and the verb-final forming intransitives, *-ê*-; the resulting stem *ohtiniskwêwê*- has the specialized meaning ‘take a wife from there, take one’s wife from there.’

Noun incorporation yields syntagmatically ordered series: the transitive stem *kanawêyim*- ‘watch over s.o., guard s.o. with care,’ when construed with the medial *-iskwêw*- ‘woman’ and the intransitive final *-ê*-, results in the incorporative verb *kanawêyimiskwêwê*- ‘watch over one’s wife/wives, carefully guard one’s wife/wives.’ It is then subject to further derivation, for instance as the transitive stem *kanawêyimiskwêwât*- ‘guard s.o. as one’s wife’; or with the noun final *-win*- forming the abstract noun *kanawêyimiskwêwêwin*- ‘watching over one’s wife/wives, carefully guarding one’s wife/wives’; or with the habitual suffix *-ski*-, giving rise to *kanawêyimiskwêwêski*- ‘habitually watch over one’s wife/wives, jealously guard one’s wife/wives at all times.’

Stems incorporating the same nominal constitute a paradigmatic set. They may not only be based on fully transitive stems, overt as in the above examples or covert as in *nâtiskwêwê*- ‘seek a wife; fetch one’s wife’ (cf. *nât*- ‘fetch s.o.’); but also on stems belonging to the paradigmatic class of mostly vowel-final stems that combines intransitive and transitive stems without marking the latter inflexionally, e.g., *mêkiskwêwê*- ‘give a woman in marriage; give (her) in marriage’ (cf. *mêki*- ‘give [it/him] away’). Finally, they may also be primary, based on a mere root instead of a full stem, e.g., *nôtiskwêwê*- ‘pursue a woman, court a woman.’

Even with intransitive stems, a medial such as *-iskwêw*- may function in an oblique relation, e.g., *âci-moskwêwê*- ‘tell about one’s affairs with women’ (cf. *âcimo*- ‘tell a story’).

Such medials also appear in completely distinct constructions: with initial and medial forming a single larger constituent construed with the final *-ê*-, such verbs mean either ‘have X,’ e.g., *oskiskwêwê*- ‘have a new wife,’ *nîsôskwêwê*- ‘have two wives,’ or ‘be X,’ e.g., *kakâmwâtiskwêwê*- ‘be a quiet woman.’

Parallel to the stative ‘be X,’ we find the denominal stem *kakâmwâtiskwêhkê*- ‘give the impression of being a quiet woman’ (with the *-êw*- of the noun

and the suffix-initial vowel contracting to *-ê*-). The incorporative stems proper, on the other hand, give rise to further transitives, e.g., *nôtiskwêwât*- ‘court s.o. as a woman’; this is true even for the oblique case: *âci-moskwêwât*- ‘tell about one’s affair with s.o. as a woman.’ But note the difference in thematic structure in *nitomiskwêwât*- ‘ask for a woman’s hand of s.o.’ (cf. *nitomiskwêwê*- ‘ask for a woman’s hand’).

Finally, the incorporative verbs may even be construed with a full patient noun phrase. The full noun often appears in a separate clause, e.g., ... *câh-cîki niwâhkômâkanak kitâcimostawin, ê-âci-moskwêwâtacik*. ‘... you told me about very close relatives of mine, telling of your affairs with them as women’; but it may also be part of the same clause, e.g., ... *ôhi iskwêwa ôhi kê-nôtiskwêwâtimiht*. ‘... the woman who was being courted.’

Verbal Art

Cree literary form liberally exploits the combined riches of inflexion, word formation, and word order. Parallel constructions, for instance, may have the verb repeated in full and both nouns in contrast position, preceding the verb: *kisêyiniwa ka-nâtâmototawêw, nôtikwêwa ka-nâtâmototawêw*. ‘They will turn to the old men (*kisêyiniwa*), they will turn to the old women (*nôtikwêwa*).’ Or the first noun may follow the verb (again repeated in full), and the second precede it, chiasmatically: *ôtê, naway ahêw kêhtê-aya, nôtikwêwa ôtê ahêw*. ‘The old people (*kêhtê-aya*) they have put over there in the background, they have put the old women (*nôtikwêwa*) over there.’ Verbless sentences, here with the personal pronoun *niya* ‘I’ and the nouns *nêhiyaw* ‘Cree,’ *nêhiyaw-iskwêw* ‘Cree woman,’ show the same rhetorical structure: *êwako ohci mitoni niya nêhiyaw, nêhiyaw-iskwêw mitoni niya*. ‘And because of that truly a Cree am I, I am truly a Cree woman.’

Both nouns may follow the verb: *pêyakosâp ihtasiwak nôsisimak, mihcêtiwak nitâniskotâpânak*. ‘My grandchildren (*nôsisimak*) number eleven, and my great-grandchildren (*nitâniskotâpânak*) are many.’ Much the same sentence (with the verbs in the changed conjunct) also shows the order of noun and verb inverted from one clause to the next: *pêyakosâp ê-ihtasicik nôsisimak, nitâniskotâpânak ê-mihcêticik*. ‘My grandchildren number eleven, and many are my great-grandchildren.’ This is the classical figure of chiasmic reversal.

The relative position of nouns and verbs can be controlled even more dramatically when the nominal element is incorporated into the verb: *misatimwak kê-nâtacik kê-nitawi-minihkwahastimwêyan ... kê-nitawi-minihkwêyâpêkinacik ôki misatimwak*. ‘when

you fetched the horses and went to do the watering of the horses . . . when you went leading the horses to be watered.’ In the second clause, instead of the noun *misatimwak* ‘horses,’ the verb (built on the transitive stem *minihkwah-* ‘make s.o. drink’) incorporates the medial *-astimw-*; the full noun precedes the verb in the first clause and follows it in the last.

Dialects, Speakers, Sources

Much more diverse than is traditionally acknowledged, the many dialects of Cree are spoken by isolated groups, derived from family hunting bands, which in many places persisted well into the 20th century. While the exact size of the speech community has not been established, Cree has far more speakers than any other indigenous language of Canada; the published numbers, however, including the 100 000 of the 2001 Census of Canada, are mere guesswork, and even the 20 000 speakers of more realistic estimates must be viewed against the backdrop of a landmass that stretches for thousands of miles across the subarctic and the northern prairies of Canada.

Cree is spoken in a chain of dialects, each remote from the next, and though many speakers control more than one dialect, distant dialects are not mutually intelligible. (The situation is roughly analogous to that which held until the end of the 18th century and the rise of the nation-state for the Romance continuum or for the complex dialectological and sociolinguistic picture presented by Alemannic [or Swiss German], the Bavarian majority language of Austria, Standard High German, Low German, and Dutch.) None of the conventional classifications of the dialects is reliable. Even the definition of the language itself is uncertain: while some limit the term ‘Cree’ to the dialects spoken between Hudson’s Bay and the Rocky Mountains, others include those spoken between Hudson’s Bay and the Atlantic coast (otherwise called Montagnais-Naskapi but also Cri de l’Est or East Cree). The examples and textual extracts in the present article represent Plains Cree as spoken on the northern prairies.

The intricate dialect situation and the high incidence of bilingualism (especially with Ojibwe, a closely related Algonquian language with a comparable degree of dialect diversity) form the backdrop for the striking case of Michif, a language distinct from both Cree and French but combining a largely French-based nominal complex with a largely Cree-based verbal system and syntax. Negation, for example, employs both the declarative *nô* (reflecting French *non*), as in *nô wîhkât mîna nîka-itostân*. (NEG-decl ever also I.will.go.there.INDEP) ‘Never again will I go there.’ and the deontic negator *kâya* (based on Cree),

as in *kâya misčêt aštâ lisel!* (NEG-deont much put.(it).IMVE salt) ‘Don’t put much salt in!’

The 19th-century grammars and dictionaries of Howse, Lacombe and Watkins are important documents, built upon in the more technical analyses of Bloomfield and Wolfart (Plains Cree), Voorhis (Western Swampy Cree), Ellis and Bêland (Eastern Swampy Cree, Moose Cree, Atikamekw) and, for the Québec and Labrador dialects of Cree-Montagnais-Naskapi, by MacKenzie, Mailhot, Martin and others. The turn of the 21st century is marked by a renewed pedagogical tradition, ideally personified in Freda Ahenakew, herself a Cree speaker, and a surge of syntactic studies pioneered by James and Dahlstrom and continued by Blain, Branigan, Brittain, Déchaine, Junker, Reinholtz, Russell, etc.

A large collection of authentic Plains Cree literature was recorded by Bloomfield in 1925 (Early Texts). Over the last third of the 20th century, a corresponding corpus of Modern Texts has been recorded and published by Wolfart and Ahenakew.

Bibliography

- Ahenakew A (2000). *âh-âyitaw isi ê-kî-kiskêyihthabkik mas-kibkiy/They knew both sides of medicine: Cree tales of curing and cursing told by Alice Ahenakew*. Edited, translated and with a glossary by Wolfart H C & Ahenakew F. Publications of the Algonquian Text Society/Collection de la société d’édition de textes algonquiens. Winnipeg: University of Manitoba Press.
- Ahenakew F (1987). *Cree language structures: a Cree approach*. Winnipeg: Pemmican Publications.
- Ahenakew F & Wolfart H C (1991). ‘The reality of Cree morpheme-boundary rules.’ *Algonquian and Iroquoian Linguistics* 16, 27–32.
- Bear G *et al.* (1992). *kôhkomînawak otâcimowiniwâwal/Our grandmothers’ lives as told in their own words*. Told by Glecia Bear *et al.* Edited and translated by Ahenakew F & Wolfart H C. Saskatoon: Fifth House Publishers.
- Blain E M (1997). Wh-constructions in Nêhiyawêwin (Plains Cree). Ph.D. diss. University of British Columbia.
- Bloomfield L (1930). *Sacred stories of the Sweet Grass Cree*. National Museum of Canada Bulletins 60.
- Bloomfield L (1934). *Plains Cree texts*. American Ethnological Society Publications 16.
- Branigan P & MacKenzie M (2003). ‘Word order variation at the left periphery in Innu-Aimûn.’ In Wolfart H C (ed.) *Papers of the Thirty-third Algonquian Conference*. Winnipeg: University of Manitoba. 110–119.
- Brittain J (1999). ‘The distribution of the conjunct verb form in Western Naskapi and related morpho-syntactic issues.’ Ph.D. diss. Memorial University of Newfoundland. [Published, *The morphosyntax of the Algonquian conjunct verb: a minimalist approach*. New York: Garland, 2001.]

- Clarke S (1982). *North-West River (Sheshâtshît) Montagnais: a grammatical sketch*. National Museum of Man, Mercury Series, Canadian Ethnology Service Paper 80.
- Dahlstrom A (1986). Plains Cree morphosyntax. Ph.D. diss. University of California, Berkeley. [Published, New York: Garland, 1991].
- Déchaine R-M (1997). 'Nominal predication in Plains Cree.' In Pentland D H (ed.) *Papers of the Twenty-eighth Algonquian Conference*. Winnipeg: University of Manitoba. 105–135.
- Drapeau L (1996). 'Conjurors: the use of evidentials in Montagnais secondhand narratives. In Nichols J D & Ogg A C (eds.) *nikotwâsik iskwâhtêm, pâskihhtëpayih! Studies in honour of H C Wolfart*. Algonquian and Iroquoian Linguistics Memoir 13, 171–194.
- Ellis C D (1971). 'Cree verb paradigms.' *International Journal of American Linguistics* 37, 76–95.
- Ellis C D (2000). *Spoken Cree/ê-ililimonânîwâhk, Level I, West Coast of James Bay*. Edmonton: University of Alberta Press. (Level II, 2004.)
- Hirose T (2000). Origins of predicates: evidence from Plains Cree. Ph.D. diss. University of British Columbia. [Published, New York, London: Routledge, 2003.]
- James D (1991). *Preverbs and the function of clauses in Moose Cree*. The 1990 Belcourt Lecture. Winnipeg: Voices of Rupert's Land.
- James D, Clarke S & MacKenzie M (2001). 'The encoding of information source in Algonquian: evidentials in Cree-Montagnais-Naskapi.' *International Journal of American Linguistics* 67, 229–263.
- Junker M-O (2000). *Quantification in East Cree and linguistic relativity*. The 1999 Belcourt Lecture. Winnipeg: Voices of Rupert's Land.
- Kâ-Nîpitêhtêw J (1998). *ana kâ-pimwêwêhahk okakêskihkêmwina/The Counselling Speeches of Jim Kâ-Nîpitêhtêw*. Edited, translated and with a glossary by Ahenakew F & Wolfart H C. Publications of the Algonquian Text Society/Collection de la société d'édition de textes algonquiens. Winnipeg: University of Manitoba Press.
- Long M A (1999). 'Adverbial and argument-doubling clauses in Cree.' M.A. thesis, University of Manitoba.
- MacKenzie M & Jancewicz B (eds.) (1994). *Naskapi lexicon/Lexique naskapi* (3 vols). Kawawachikamach, Quebec: Naskapi Development Corporation/Société du développement des Naskapis.
- MacKenzie M et al. (eds.) (1987). *Cree lexicon: Eastern James Bay dialects*. Val d'Or: Cree School Board/Commission scolaire crie.
- Mailhot J & Lescop K (1977). *Lexique montagnais-français du dialecte de Schefferville, Sept-Iles et Maliotenam*. Direction de l'Inventaire des biens culturelles, Dossier 29. Direction générale du patrimoine, Ministère des affaires culturelles, Québec.
- Martin P (1991). *Le montagnais, langue algonquienne du Québec*. Langues et Sociétés d'Amérique traditionnelle 3. SELAF 328. Paris: Peeters.
- Minde E (1997). *kwayask ê-kî-pê-kiskinowâpahtihicik/ Their example showed me the way: a Cree woman's life shaped by two cultures*. Told by Minde E. Edited, translated and with a glossary by Ahenakew F & Wolfart H C. Edmonton: University of Alberta Press.
- Pentland D H & Wolfart H C (1982). *Bibliography of Algonquian linguistics*. Winnipeg: University of Manitoba Press.
- Reinholtz C (2002). 'On the characterization of the Cree question particle.' In Wolfart H C (ed.) *Papers of the Thirty-third Algonquian Conference*. Winnipeg: University of Manitoba. 378–391.
- Reinholtz C & Russell K (1995). 'Quantified NPs in pronominal argument languages: evidence from Swampy Cree.' In Beckman J N (ed.) *Proceedings of the Twenty-fifth Meeting of the North East Linguistic Society* 1. 389–404.
- Russell K (1999). 'The "Word" in two polysynthetic languages.' In Hall T A & Kleinhenz U (eds.) *Studies on the phonological word*. Amsterdam: J. Benjamins. 203–221.
- Russell K & Reinholtz C (1996). 'Hierarchical structure in a non-configurational language: asymmetries in Swampy Cree.' In Camacho J (ed.) *Proceedings of the fourteenth west coast conference in formal linguistics*. 431–445.
- Scott S et al. (1995). *âtalôhkâna nêsta tipâcimôwina/Cree legends and narratives from the west coast of James Bay: text and translation*. Told by Scott S et al., edited and with a glossary by Ellis C D. Publications of the Algonquian Text Society/Collection de la société d'édition de textes algonquiens. Winnipeg: University of Manitoba Press.
- Vandall P & Douquette J (1987). *wâskabikaniwiyiniw-âcimowina/Stories of the House people*. Told by Vandall P & Douquette J. Edited, translated and with a glossary by Ahenakew F. Publications of the Algonquian Text Society/Collection de la société d'édition de textes algonquiens. Winnipeg: University of Manitoba Press.
- Voorhis P H et al. (1977). *A Cree phrase book based on the dialects of Manitoba*. Brandon, Manitoba: Department of Native Studies, Brandon University.
- Whitcalf S (1993). *kinêhiyâwîwininaw nêhiyawêwin/The Cree language is our identity: the La Ronge lectures of Sarah Whitcalf*. Edited, translated and with a glossary by Wolfart H C & Ahenakew F. Publications of the Algonquian Text Society/Collection de la société d'édition de textes algonquiens. Winnipeg: University of Manitoba Press.
- Wolfart H C (1973). *Plains Cree: a grammatical study*. American Philosophical Society, Transactions, n.s., vol. 63(5). Philadelphia.
- Wolfart H C (1996). 'Sketch of Cree, an Algonquian language.' In Goddard I (ed.) *Languages*. 390–439. Sturtevant W C (gen. ed.) *Handbook of North American Indians*, vol. 17. Washington, DC: Smithsonian Institution.
- Wolfart H C (2002). 'Cree literature.' In New W H (ed.) *Encyclopedia of literature in Canada*. Toronto: University of Toronto Press. 243–247.
- Wolfart H C (2005). 'Cri.' In Queixalós F (ed.) *Amérique*. Bonvini E & Peyraube S (gen. eds.) *Dictionnaire des*

langues du monde. Auroux S (gen. ed.) *Encyclopédie des sciences du langage*. Paris: Presses Universitaires de France. Wolfart H C & Ahenakew F (1998). *The student's dictionary of literary Plains Cree, based on contemporary texts*. Algonquian and Iroquoian Linguistics, Memoir 15. Winnipeg.

Wolfart H C & Carroll J F (1981). *Meet Cree: a guide to the Cree language* (2nd edn.). Edmonton: University of Alberta Press/Lincoln, London: University of Nebraska Press.

Creek

K M Booker, The University of Kansas, Lawrence, KS, USA

© 2006 Elsevier Ltd. All rights reserved.

Historical Background

When encountered by European explorers in the 16th century, the ancestors of the Native Americans belonging to what is now called the Creek and Seminole nations were living along the rivers of the present-day states of Alabama and Georgia. The English called the natives 'Creeks' because of the settlement patterns along the rivers, but these Native Americans called themselves *este maskoke* /isti ma(:)sko:ki/ 'the Muskogee people' and referred to their language as *este maskoke empunvka* /isti ma(:)sko:ki imponaka/ 'the language of the Muskogee people.' Conflicts with European settlers resulted in the removal of the Muskogee people (Creeks) from their ancestral lands. Some of these people joined with Hitchiti and Mikasuki speakers and fled into the swamps of Florida, becoming the 'Seminoles,' a term derived from Spanish *cimarrón* 'wild, untamed.' Others were forcibly moved to Indian Territory (the present state of Oklahoma) in the 1830s. After the Seminole Wars, a Florida contingent also settled in Indian Territory. These two Creek-speaking groups separated into the Creek (Muskogee) and Seminole nations. In Florida, the term 'Seminole' has come to refer to the political unit composed of Mikasuki and Seminole (Creek) speakers. The various population dispersals have resulted in three separate Creek dialects – Muskogee and Oklahoma Seminole, spoken by some 5000 descendants of those who settled in Oklahoma (Hardy, 2005), and Florida Seminole, spoken by fewer than 100 Creek speakers residing in Florida. The dialects differ mainly in vocabulary and are mutually intelligible (Martin and Mauldin, 2000).

On settling in Indian Territory, the Creeks were heavily missionized. With the help of educated native speakers, an alphabet was devised that is still in use today. Numerous religious publications, even a

newspaper, were published in the language. Generations of Creeks learned to read and write their language using the traditional alphabet, and it remains a vital part of Creek culture and ethnic identity. The Creek language belongs to the Muskogean family along with the extant languages Choctaw, Chickasaw, Alabama, Koasati, and Mikasuki. The interrelationships among the various branches are a topic of current research.

Phonology

The traditional Creek alphabet is a semiphonemic representation of the language that does not indicate pitch. For accuracy, the data presented here are in phonemic notation. Because instructional materials use the Creek alphabet, the traditional symbols are enclosed in parentheses to better illustrate the correspondences between the two systems.

Creek has 13 consonants. The three stops and one affricate, *p*(*p*), *t*(*t*), *c* [tʃ] (*c*), and *k*(*k*), are articulated at the labial, alveolar, alveopalatal, and velar positions, respectively. They are lenis, unaspirated, and are voiced between vowels. There are four fricatives: *f*(*f*), which is bilabial in some speakers but labiodental in others; *ɬ*(*r*), a voiceless alveolar lateral; *s*(*s*), an alveolar sibilant that may be retroflexed in some speakers; and the glottal *h*(*h*). Resonants include two nasals, bilabial *m*(*m*) and alveolar *n*(*n*); the alveolar lateral *l*(*l*); and two glides, alveopalatal *y*(*y*, *e*) and velar *w*(*w*, *u*, *o*). The vowel system consists of three phonemic short vowels, *i*(*e*), *a*(*a*, *ʌ*), and *o*(*o*, *u*), contrasting with three corresponding long vowels, *i:*(*ē*, *e*), *a:*(*a*), and *o:*(*o*), and the diphthongs *ey*(*i*), *aw*(*ʌo*), and *oy*(*ue*).

Creek has a pitch accent system with three contrasting tones, high /'/, falling /'/, and extra high /'/. High pitch is primarily nonphonemic, with iambic assignment based on syllable structure. It is phonemic (fixed), however, in some lexemes and a few grammatical morphemes. Falling tone is both lexical and grammatical. Extra high pitch occurs only in the intensive stem grade (EGR; discussed later). In words

with more than one accented syllable, a downdrift phenomenon occurs in which an accented syllable is one step lower in pitch than the accented syllable it follows (see Haas (1977) and Martin and Johnson (2002) for the specifics of Creek pitch).

Morphology

Creek is a largely agglutinating language with minimal noun morphology and extensive verb morphology. Prefixation, suffixation, infixation, vowel lengthening, tonal accent, and suppletion are used to mark grammatical categories.

Noun Morphology

Nominal case marking occurs only on indefinite nouns, *-t* for the nominative and *-n* for the oblique case. Definiteness is indicated by the lack of suffixation. Compare Examples (1) and (2), in which ‘snake’ is indefinite in both examples, whereas ‘dog’ is indefinite in Example (1) but definite in Example (2):

- (1) citto-t ifa-n ákkis
snake-SUBJ dog-OBL bit
‘A snake bit a dog.’
- (2) citto-t ifa ákkis
snake-SUBJ dog bit
‘A snake bit the dog.’

Most nouns are unmarked for number, although Creek does have three lexically determined plural suffixes restricted to human referents: the collective *-álki* in *masko:k-álki* ‘Creek people,’ *-aki* in *acol-aki* ‘old men,’ and *-ta:ki* in *hopoy-ta:ki* ‘children.’

Possession is indicated by one of two sets of nominal prefixes. Possessed nouns intimately associated with the possessor, such as most body parts and kin terms, take the inalienable (Type II) prefixes, as in *ca-cokwa* ‘my mouth’ and *ca-posi* ‘my grandmother.’ Nouns having a looser relationship with the possessor take the alienable (Type III) set of possessive prefixes, such as *an-coko* ‘my house’ and *an-hissi* ‘my friend.’

Derivational morphology includes the productive diminutive *-oci* and augmentative *-takko* (from *tákki*: ‘big’) suffixes (compare *ifa* ‘dog’ with *if-oci* ‘puppy’ and *ico* ‘deer’ with *co-takko* ‘horse’).

Several derivational processes create nouns from verbs. Agentives are formed by suffixing *-a* to the verb stem with concomitant vowel lengthening: *ali:kc-a* ‘a doctor’ is formed from *alík-ita* ‘to cure.’ The instrumental prefix *is(s)-* combined with the nominalizer *-ka* creates derived nouns, as in *is-lá:f-ka* ‘knife’ (*laff-ita* ‘to cut with a knife’). Other nouns are formed with the instrumental and the infinitive, as in *is-łolak-ita* ‘awl’ (*łolak-ita* ‘to pierce’).

Verb Morphology

Creek has a complex system of inflectional verb morphology utilizing prefixes, suffixes, infixes, pitch, and suppletion to mark person, tense, aspect, and number. Although independent pronouns exist, they are used only for emphasis. The verb is obligatorily marked for person with one of two sets of subject affixes. Transitive and intransitive verbs in which an actor exercises control over the event are marked with the Type I suffixes: *-eyl-ay* ‘I,’ *-íck* ‘you.SG,’ *-i:/-iy* ‘we,’ *-á:ck/ -á:cc* ‘you.PL.’ Third person is unmarked. The sentences in Examples (3a)–(3c) illustrate the conjugation of *hic-ita* ‘to see’ in the lengthened grade (LGR; see later) with the Type I suffixes:

- (3a) hi:c-ey-s hi:c-i:-s
see.LGR-I-DECL see.LGR-we-DECL
‘I see.’ ‘We see.’
- (3b) hi:c-íck-is hi:c-á:ck-is
see.LGR-you.SG-DECL see.LGR-you.PL-DECL
‘You see.’ ‘You see.’
- (3c) hi:c-is hi:c-is
see.LGR-DECL see.LGR-DECL
‘He sees.’ ‘They see.’

With stative verbs, the Type II prefixes are used: *ca-* ‘I,’ *ci-* ‘you,’ *po-* ‘we.’ The third person is unmarked and there is no number distinction in the second person. These are the same prefixes that are used to mark inalienable possession on nouns. Examples (4a)–(4c) illustrate the conjugation of the stative verb *má:h-i*: ‘tall’:

- (4a) ca-má:h-i:-s po-má:h-i:-s
I-tall-STATE-DECL we-tall-STATE-DECL
‘I’m tall.’ ‘We’re tall.’
- (4b) ci-má:h-i:-s ci-má:h-i:-s
you-tall-STATE-DECL you-tall-STATE-DECL
‘You’re tall.’ ‘You’re tall.’
- (4c) má:h-i:-s má:h-i:-s
tall-STATE-DECL tall-STATE-DECL
‘He’s tall.’ ‘They’re tall.’

A small set of intransitive verbs take either set of affixes. Note the difference in meaning between Examples (5) and (6):

- (5) nockihl-ey-s
sleepy.HGR-I-DECL
‘I fell asleep.’
- (6) ca-nockil-i:-s
I-sleepy-STATE-DECL
‘I’m sleepy.’

The Type II prefixes also index direct objects, as in Example (7):

- (7) ca-híhc-is
me-see.HGR-DECL
‘He saw me.’

A third set of pronominal affixes is used to mark dative objects; they are the same as the prefixes used to encode alienable possession:

- (8) in-háhy-ey-s
for.him-make.HGR-I-DECL
'I made it for him.'

Other pronominal verbal prefixes include the reflexive *i:-* 'one's self' (Example (9)) and the reciprocal *ti-* 'each other' (Example (10)):

- (9) i:-nókł-eyc-s
REFL-burn-CAUS-DECL
'He burned himself.'

- (10) ti-pak-ita
RECIP-join-INF
'to get married.'

The verb stem undergoes changes in pitch, vowel lengthening, and infixation, resulting in five aspectual ablaut grades. The zero grade is the unaltered stem, as found in the infinitive *hic-ita* 'to see.' In the lengthened grade (LGR), the stem vowel is lengthened (Example (3)), usually indicating the continuative aspect. The falling tone grade (FGR) encodes the resultative aspect, a state resulting from the action of the verb (see later, Examples (24) and (25)). The extra high pitch grade (EGR) involves nasal infixation (indicated by ⁿ), extra long vowel lengthening, and extra high pitch, signaling the intensive aspect. Examples (11a)–(11c) illustrate three degrees of smallness, the first with the unmodified verb, the second with the intensive morpheme, and the third with the intensive morpheme plus the EGR:

- (11a) cótk-i:-s
small.SG-STATE-DECL
'It's small.'
- (11b) cótk-ós-i:-s
small.SG-INTENS-STATE-DECL
'It's very small.'
- (11c) cõ::ⁿtk-ós-i:-s
small.SG.EGR-INTENS-STATE-DECL
'It's really, really small.'

The final stem grade, the *b*-grade (HGR), involves infixing an *h* with high pitch on the preceding stem vowel, as in Examples (5), (7), and (8). The HGR has multiple uses, the most common of which is to indicate an instantaneous or an immediate past action.

Creek has five past tenses marked by suffixes and stem change. Past I, the immediate past, marks events in the recent past, i.e., earlier today or last night. It is realized by using the HGR of the stem. Examples (5) and (8) showed one allomorph of the HGR. Past II, the recent past, refers to events happening up to a year ago. The tense suffix is *-ánk* and it cooccurs with the LGR, as in Example (12):

- (12) hi:c-ay-ánk-s
see.LGR-I-PASTII-DECL
'I saw it (a while ago).'

Events occurring a year or several years ago take the intermediate past suffix *-(i)mát* (Past III) with the verb stem in the LGR, as shown in Example (13):

- (13) hi:c-ey-mát-s
see.LGR-I-PASTIII-DECL
'I saw it (long ago).'

The remote past tense suffix *-ánta* (Past IV) marks events happening a long, long time ago; the verb stem is in the LGR, as in Example (14):

- (14) hi:c-ay-ánta-s
see.LGR-I-PASTIV-DECL
'I saw it (long, long ago).'

Past V, the indefinite past, is reserved for events outside the speaker's sphere of reference. It is rarely used in conversation but is always found in traditional folktales. In Example (15), the tense suffix is *-ati:*; the stem vowel is not lengthened because it is in a heavy syllable:

- (15) honanwa acól-i:-t leyk-ati:-s
man old-STATIVE-INDEF sit-PASTV-DECL
'Once upon a time there was an old man.'

Future time is marked by the future suffix *-áti:* in Example (16), and the intensitive *-áha:n* in Example (17):

- (16) ay-áti:-s
go.SG-FUT-DECL
'He will go.'
- (17) ay-áha:n-is
go.SG-INTENT-DECL
'He's going.'

Creek marks the declarative, interrogative, and imperative modes, all with verb-final suffixes. The declarative *-(i)s* is exemplified in the preceding examples. The interrogative *-a* occurs on yes/no questions, as in Example (18), whereas *-a:* is the final suffix in questions seeking information, as in Example (19):

- (18) níhs-ick-a
buy.HGR-you.SG-INTERROG
'Did you buy it?'
- (19) nâ:ki-n níhs-ick-a:
thing-OBL buy.HGR-you.SG-INTERROG
'What did you buy?'

The affirmative imperative is formed with *-as* in the singular and *-aks* in the plural, e.g., *pap-as* 'eat it' (to one person) and *homp-aks* 'eat' (to more than one person). For the hortative, the suffix *-ik* is added before the imperative ending, as in *lît-ik-as* 'let him run.'

Negative statements are formed with *-iko* suffixed to active verbs (Example (20)) and *-iko:* suffixed to stative verbs (Example (21)):

- (20) nâ:k-n nis-íko-t †-aláhk-is
thing-OBL buy-NEG-SS went.and-arrive.here.
HGR-DECL
'He went and came back without having bought anything.'
- (21) satałakko ci-yá:c-íko:t ô:m-a
apple you-like-NEG-SS be.FGR-INTERROG
'You don't like apples?'

Derivational verb morphology includes instrumental and locative prefixes and a causative suffix. The instrumental *is(s)-* ‘with’ is an applicative suffix that adds an additional argument to the verb, as in Example (22):

- (22) iss-âla:k-is
INSTR-come.back.SG.LGR-DECL
'He's coming back with it.'

Creek has a number of locative verbal prefixes. Three mark the static position of a referent with respect to a particular plane, usually the ground; *oh(h)-* in Example (23) and *tak(k)-* in Example (24) locate their referents with respect to the ground. In Example (25), however, the plane of reference is the table. The locative *ak(k)-* is always used to refer to objects situated in the water.

- (23) oh-léyc-as
on.above-set-IMP
'Put it on it (e.g., a table).'
- (24) takk-apô:k-i:s
on.ground-sit.PL.FGR-we-DECL
'We're seated on the ground.'
- (25) sampa-t ohhompita lica-n ak-léyk-is
basket- table under- down.low-
SUBJ OBL sit.FGR-DECL
'A basket's sitting under the table.'

Other locative prefixes refer to motion, as in *(i)ʔ* ‘arrive over there and ...’ in Example (26) and *y(i)-* ‘arrive here and ...’ in Example (27). The prefix *acok(h)-* specifies motion toward the speaker, with no implied arrival (Example (28)).

- (26) 4-in-tac-ácc-as
go.and-for.her-cut-you.PL-IMP
'Go over there and cut it for her.'
- (27) yi-náfk-as
come.and-hit.INSTANT-IMP
'Come here and hit it!'
- (28) acókh-a:t-is
towards.speaker-come.SG.LGR-DECL
'He's coming towards me.'

The causative suffix *-ic* combines with an underlying verb-final vowel to produce the allomorphs *-eyc* (*a + ic*), *-i:c* (*i + ic*), *-oy* (*o + ic*), and *-ic* (*C + ic*), e.g., *il-ita* ‘one to die’ and *il-i:c-ta* ‘to kill one.’

Although the Creek language rarely distinguishes singular and plural nouns, several Creek verbs mark number with suppletive forms, i.e., there are different verb roots for singular vs. plural or singular, dual, and plural. The majority of verbs that supplete are positionals (e.g., ‘sit,’ ‘stand,’ and ‘lie’) or verbs of motion (e.g., ‘come’ and ‘go’). Intransitive suppletive verbs indicate subject number. The verb ‘sit’ has three forms, *leyk-ita* ‘one to sit,’ *ka:k-ita* ‘two to sit,’ and *apo:k-ita* ‘three or more to sit,’ as does the verb of motion, ‘run’: *litk-ita* ‘one to run,’ *toko!k-ita* ‘two to run,’ and *pifa:tk-ita* ‘three or more to run.’ Transitive suppletive verbs mark object number, as in *is-ita* ‘to take one’ and *caw-ita* ‘to take more than one.’ A small number of nonlocative verbs have suppletive roots as well, such as *il-ita* ‘one to die,’ as opposed to *pasatk-ita* ‘more than one to die.’ One stative verb is suppletive, i.e., *cótk-i*: ‘small.SG’ and *lopóck-i*: ‘small.PL.’ Still other verbs differentiate number by suppletion, affixation, and reduplication in various combinations.

Syntax

A Creek sentence consists of an inflected verb. With overt noun phrases, the order of constituents is subject-object-verb, as was shown in Examples (1) and (2). Rarely, however, does more than one overt noun occur in a clause. Modifiers follow the noun (*ico hámkin* (deer one) ‘one deer’). Postpositions follow their heads, as in *lica* ‘under’ (see Example (25)).

Constructions with an auxiliary verb are common (see Example (21)), such that the auxiliary *om-ita* ‘to be’ follows the main verb. Certain verbs of motion, such as *wilak-ita* ‘two to go about’ in Example (29), function both as independent verbs and auxiliaries. Clause participants are indexed by a switch reference system, whereby a clause-final suffix indicates whether the subject of the following clause is the same (SS) or different (DS) from the one in the marked clause. The English translation in Example (29) is opaque without the subscript numbers to track the participants, but the Creek sentence is clear because of the different subject suffix:

- (29) nafêyk-in a:-nafêyk-in s-tak-wila:k-ey-s
hit.HGR-DS away-hit.HGR-DS INSTR-on.ground-
go.about.DU.
LGR-PASTI-DECL
'He₁ hit him₂, and he₂ hit him₁ back with it.'

For an example of the same subject suffix, see Example (20).

For Further Study

Two published Creek grammars are available for more in-depth language study. Hardy (2005) contains a text with an extensive linguistic sketch. Innes *et al.* (2004) is the first in an anticipated series of pedagogical texts. A comprehensive dictionary by Martin and Mauldin (2000) has entries in the traditional alphabet, with transliterations in phonemic transcription. A volume of traditional folktales by Martin *et al.* (2004) has the Creek text and free English translation in parallel columns (for earlier works on Creek, see Booker (1991)).

Bibliography

- Booker K M (1984). 'Directional prefixes in Creek.' In Rood D S (ed.) *Proceedings of the 1983 Mid-America Linguistics Conference*. Boulder: University of Colorado. 59–87.
- Booker K M (1991). *Languages of the aboriginal southeast: an annotated bibliography*. Metuchen and London: Scarecrow Press.
- Booker K M (1993). 'Nasalization and question formation in Creek.' In Smith E & Zéphir F (eds.) *Proceedings of the 1992 Mid-America Linguistics Conference*. Columbia: University of Missouri. 373–388.
- Haas M R (1940). 'Ablaut and its function in Muskogee.' *Language* 16, 141–150.
- Haas M R (1977). 'Tonal accent in Creek.' In Hymes L M (ed.) *Southern California papers in linguistics 4: studies in stress and accent*. Los Angeles: University of Southern California. 195–208.
- Hardy D E (2005). 'Creek.' In Hardy H & Scancarelli J (eds.) *Native languages of the southeastern United States. Studies in the anthropology of North American Indians*. Lincoln and London: University of Nebraska Press. 200–245.
- Innes P, Alexander L & Tilkens B (2004). *Beginning Creek: Muskoke emponvkv*. Norman: University of Oklahoma Press.
- Martin J B (1991). 'Lexical and syntactic aspects of Creek causatives.' *International Journal of American Linguistics* 57, 194–229.
- Martin J B & Johnson K (2002). 'An acoustic study of tonal accent in Creek.' *International Journal of American Linguistics* 68, 28–50.
- Martin J B & Mauldin M M (2000). *A dictionary of Creek/Muskogee: with notes on the Florida and Oklahoma Seminole dialects of Creek*. Lincoln and London: University of Nebraska Press.
- Martin J B, Mauldin M M & McGirt J (eds./transl.) (2004). *Totkv mocvse/new fire: Creek folktales by Earnest Gouge*. Norman: University of Oklahoma Press.

Crow

R Graczyk, St. Charles Mission, Pryor, MT, USA

© 2006 Elsevier Ltd. All rights reserved.

Location, Speakers, and External Relationships

The Crow, or *Apsáalooke*, language is spoken primarily on and near the Crow reservation in southeastern Montana. There are over 4000 speakers, most of whom are adults, although there are a few children who still speak the language, and many more who understand it. For most adults, Crow is still the language of the home and the preferred language for interaction with other tribal members.

Crow, along with Hidatsa, is a member of the Missouri River subgroup of the Siouan language family (see **Siouan Languages**). Crow and Hidatsa have diverged considerably from the other languages of the family, suggesting that this subgroup may have been the first to separate from the protolanguage.

Orthography and Phonology

Crow is written in a practical orthography developed by the Crow Agency Bilingual Education Program and the Wycliffe Bible translators in the late 1960s. The values of the letters are roughly as in English, with the following exceptions: *ch* = /č/, *sh* = /š/, *tch* = /čč/, *ssh* = /šš/, *sch* = /šč/, ? = glottal stop, and *x* represents the velar fricative. Long vowels are written as digraphs: *aa*, *ee*, etc.

The consonant sounds of Crow are given in Table 1. The voiced sonorants *m* and *n* have three allophones: *w* and *l* between vowels, *b* and *d* word-initially and following an obstruent, and *m* and *n* elsewhere;

Table 1 Crow consonant inventory

| Consonant | Labial | Alveolar | Palatal | Velar | Glottal |
|------------|--------|----------|---------|-------|---------|
| Stops | p | t | ch | k | (?) |
| Fricatives | | s | sh | x | |
| Sonorants | m | n | | | h |

Table 2 Crow vowel inventory

| Vowel | –Round | | +Round | |
|------------|---------|-------|--------|-------|
| | Long | Short | Long | Short |
| High | ii | i | uu | u |
| Mid | ee | | oo | |
| Low | aa | a | | |
| Diphthongs | ia (ea) | | ua | |

m occurs in free variation with *b* word-initially, although *b* is the more common realization. The glottal stop (written with the question mark) is a defective phoneme that occurs only as the sentence-final marker of an interrogative. Crow has a single stop series. Stops are aspirated word-initially and -finally, and as the second member of a cluster. Geminate stops are treated as clusters and are aspirated. Single stops between vowels are lax, unaspirated, and often voiced. The *k* is palatalized after *i*, *ih*, *e*, *eh*, *ch*, and *sh*. Fricatives are lax (and sometimes voiced) intervocalically.

The vowel inventory of Crow is given in Table 2. The diphthongs are realized as long vowels followed by an off-glide; *ea* is a marginal diphthong that occurs in only two stems. Crow and Hidatsa lack the nasalized vowels found in other Siouan languages. Length is phonemic in Crow (e.g., *báalaa* ‘winter’, *bálaa* ‘money’) with the exception of the mid vowels, which are always phonemically long, although users of the practical orthography often spell words with short *e* and *o*. Crow has a pitch-accent system that contrasts long falling, long high, and short accented vowels (*chúisa* ‘tail’, *chii* ‘pack on back’, *axíchi* ‘wet’). Long vowels preceding the accent are high in pitch, and all vowels following the accent have low pitch. Short vowels that occur between a high vowel and the accent assimilate to high pitch. The accent may occur on any syllable of the word. Accent is distinctive in Crow: there are minimal pairs that differ only in the placement of the accent, as in *húupa* ‘handle’ and *huupá* ‘shoe’.

Morphology

Nouns are inflected for possessor with two different inflectional patterns, one for alienable possession and the other for inalienable possession (kin terms, body parts, and a few other objects closely associated with a person, such as items of clothing). The following examples demonstrate alienable and inalienable paradigms:

| Alienable | Inalienable |
|----------------------------|-------------------|
| bas-búupche ‘my ball(s)’ | b-apé ‘my nose’ |
| dis-buupche ‘your ball(s)’ | d-ápe ‘your nose’ |

| | |
|-------------------------------|----------------------------|
| is-búupche ‘his/her balls(s)’ | ø-apé ‘his/her nose’ |
| bas-búupt-uua ‘our ball(s)’ | b-ap-úua ‘our noses’ |
| dis-buupt-uua ‘your ball(s)’ | d-áp-uua ‘your (PL) noses’ |
| is-búup-tuua ‘their ball(s)’ | ø-ap-úua ‘their noses’ |

There are several other inflectional paradigms for inalienably possessed nouns. The plural marker (*uua* in these examples) on a possessed noun indicates that the possessor is plural; the forms are ambiguous as to whether the possessum is singular or plural. Crow has a series of articles that are suffixed to the final word of a noun phrase: *iichíil-eesh* ‘the horse’ (definite), *iichíili-m* ‘a horse’ (indefinite specific), and *iichíil-eem* (indefinite nonspecific). Plural number for both nouns and verbs is marked by a suffix: *uu(a)* after short vowels and *o* or *u* after long vowels.

Verbal morphology is considerably more elaborated in Crow. Crow verbs are inflected according to an active/stative pattern, with the subjects of some intransitive verbs utilizing the same pronominal prefixes as the objects of transitive verbs:

| Object of transitive active | Subject of stative |
|---|-----------------------------------|
| bii-lichík ‘he hit me’ | bii-waakupáak ‘I am sick’ |
| dii-lichík ‘he hit you’ | dii-waakupáak ‘you are sick’ |
| ø-ø-dichík ‘he hit him/her’ | ø-baakupáak ‘she/he is sick’ |
| balee-lichík ‘he hit us’ | balee-waakupáak ‘we are sick’ |
| dii-lit-úuk ‘he hit you all/they hit you all’ | dii-waakupáauk ‘you all are sick’ |
| ø-ø-dit-úuk ‘they hit/him/her/them’ | ø-baakupáauk ‘they are sick’ |

Both active and stative verbs lack an overt third-person pronominal prefix for either subject or object. Other pronominal prefixes mark the subjects of active verbs, both transitive and intransitive:

| Intransitive active | Transitive active |
|----------------------------|--|
| baa-xalússhik ‘I run’ | dii-wah-kuxshík ‘I helped you’ |
| da-xalússhik ‘you run’ | bii-láh-kuxshik ‘you helped me’ |
| ø-xalússhik ‘she/he runs’ | ø-ø-kuxshík ‘she/he helped them’ |
| baa-xalússuuk ‘we run’ | dii-wah-kuxsúuk ‘I helped you all’/‘we helped you’ |
| da-xalússuuk ‘you all run’ | bale-láh-kuxsuuk ‘you all helped us’ |
| ø-xalússuuk ‘they run’ | bale-ø-kuxsúuk ‘they help us’ |

Object prefixes ordinarily precede the subject prefixes. With second-person objects, there can be ambiguity as to whether subject or object is singular or plural.

Crow has a number of different inflectional patterns for active verbs that have arisen from the combination of pronominal prefixes with locative and instrumental prefixes. The following examples demonstrate these paradigms:

| | | | |
|-------------------|-----------------|--------------|--------------------|
| du(u)- 'by hand' | 1.sg-bulusshuak | 2-dilússhuak | 3-dússhuak 'bend' |
| da(a)- 'by mouth' | 1.sg-balapxík | 2-dalápxik | 3-dáapxik 'bite' |
| ala- 'by foot' | 1.sg-baatshik | 2-dáatshik | 3-alatshik 'slip' |
| pa- 'by pushing' | 1.sg-bapchilek | 2-dápchilek | 3-páachilek 'push' |

A variety of suffixes may follow the verb stem. Some encode aspectual notions, as in *áhi* 'punctual, instantaneous' and *i* 'habitual'. Others function as manner adverbials; examples are *aachíllíhí* 'rather, like' (approximative), *aahi* 'here and there' (distributive), *kaáshi* 'very much, really' (augmentative), and *káata* 'little' (diminutive). Reduplication of a portion of the stem adds a distributive or augmentative sense, as in *dappaxí* 'split', *dappáppaxii* 'chop into little pieces' and *ihchipúá* 'jump', *ihchipúpuahi* 'jump up and down'.

Morphosyntax

Although Crow is a head-marking subject-object-verb language, overt noun phrases need not be present to constitute a grammatical sentence; the verb and its pronominal prefixes are sufficient. Crow morphosyntax tends strongly toward polysynthesis and incorporation; as a result, a sentence often consists of a single morphologically complex phonological word:

baa-w-aash-baa-lée-wia-waa-ssaa-k
INDEF-1-hunt-1-go-want.to-1-NEG-DECL
'I'm not going to go hunting'.

This sentence consists of three verbs: *aashí* 'hunt', *dée* 'go', and *wiá* 'want to' (auxiliary verb). Each of the verbs is marked for person of subject. *Baa* marks the object of 'hunt' as indefinite. The final declarative marker *k* is preceded by the negative suffix *ssaa*.

Nominal object incorporation is also common when referring to habitual activities, with the incorporated object preceding the verb, as in *ílii-laxxoxxi* 'peel teepee poles', *báláa-kaali* 'ask for money', *iichíil-aakinnee* 'ride horseback', and *bil-issbíi* 'drink water'. When transitive verbs lack a specific object, *baa* (indefinite object) is prefixed to the verb stem, as in *baa-kaali* 'ask for (things)' and *baa-issbíi* 'drink, be drinking'.

Crow sentences end with one of a series of final markers of utterance type. The principal ones are *k* (declarative), *h* (imperative), and ? (interrogative):

| | |
|-----------------------------|---|
| Baáhpúuo kuss- baa-lée-k | 'I went to Pryor' (declarative) |
| Baáhpúuo kuss-da- lée-? | 'Did you go to Pryor?' (interrogative) |
| Baáhpúuo kuss- dáa-h | 'Go to Pryor!' (imperative) |

Other sentence-final markers code evidentiality:

| | |
|---------------------------|---|
| Baáhpúuo kuss- dée-sho | 'He must have gone to Pryor' (indirect evidence) |
| Baáhpúuo kuss- dée-wis | 'She probably went to Pryor' (probability) |

Subordinate clauses are marked by a clause-final suffix, as in *huu-lák* 'if he comes', *baakuhpáa-lasshen* 'because she is sick', and *xaláa-lahtaa* 'even if it rains'.

Crow marks switch-reference in nonindependent clauses: clause-final *ak* indicates that the subject of the following clause will be the same, whereas *m* indicates that the subject of the following clause will be different. It is not unusual to find long chains of clauses linked by *ak* and *m*, with only the final clause in the series terminating in declarative *k*. In noun phrases, demonstratives precede nouns, and other modifiers follow. The determiner is phrase-final, as in *hinne bachée-sh* 'this man-DEF'. Relative clauses are internally headed, and the head noun is marked with the indefinite specific determiner *m*. The agentive subject of the relative clause is marked by the relativizer (REL) *ak*:

| | | |
|-------------------------------------|--------------|---------------|
| [hileen bachée-m | ak-húua-sh] | aw-ák-uu-k |
| these men-INDEF | REL-come-DEF | 1-see-PL-DECL |
| 'We saw these men who were coming'. | | |

Crow has a limited set of postpositional suffixes that combine with nouns, demonstratives, or other postpositions to form complex postpositions: *n* 'at, in, on' (location), *ss(aa)*, *ss(ee)* 'to, toward' (goal), *taa* 'along' (path), and *kaa* 'from' (source):

ótchia héelapee-n
night middle -at
'in the middle of the night'

bin-náaskee-taa
water-edge-along
'along the water's edge'

awaxaawé ku-ssee
mountain it -toward
'toward the mountain'

Under certain conditions, postpositions may be incorporated by the following verb:

| | |
|-------------------------|---------------|
| Ammalapáshkua | ku-ss-dée-k |
| Billings | it-to-go-DECL |
| 'She went to Billings'. | |

Bibliography

- Graczyk R (1987). 'Switch reference in Crow: an emergent category.' In Kroeber P D & Moore R E (eds.) *Native American languages and grammatical typology*. Bloomington: Indiana University Linguistics Club. 119–135.
- Graczyk R (2005). *Apsáalooke aliláau: a grammar of Crow*. Lincoln, NE: University of Nebraska Press.
- Kaschube D V (1967). *Structural elements of the language of the Crow Indians of Montana*. University of Colorado studies, series in anthropology 14. Boulder: University of Colorado Press.
- Kaschube D V (1978). *Crow texts*. IJAL Native American texts series, monograph 2. Chicago: University of Chicago Press.
- Lowie R H (1960). *Crow texts*. Berkeley: University of California Press.
- Lowie R H (1960). *Crow word lists*. Berkeley: University of California Press.
- Medicine Horse M H (1987). *A dictionary of everyday Crow*. Crow Agency, MT: Bilingual Materials Development Center.
- Read J A S (1978). *A sociolinguistic study of Crow language maintenance*. Dissertation: University of New Mexico, Albuquerque.

Cupeño

J H Hill, University of Arizona, Tucson, AZ, USA

© 2006 Elsevier Ltd. All rights reserved.

Cupeño is a Uto-Aztecan language in the Cupan group of the Takic subfamily. Until 1903, speakers lived primarily in two villages, Kupa and Wilakalpa, located at hot springs sites southeast of Mount Palomar in southern California. In 1903, the Cupeño were forced from their homes by the legal machinations of White landowners (Hyer, 2001). Many live today at Pala, California. The Cupeño continue to maintain close economic, social, and ritual ties with speakers of other Takic languages, including Luiseño and Cahuilla, and with speakers of the Iipay (Diegueño) languages, in the Yuman family. These connections are reflected in loan material in Cupeño (Hinton, 1991), including the name of their principal village, *Kupa*, from Iipay Aa *ha-kupin* 'water-warm.' No fluent first-language speakers of the language remain, but the language is studied and is used in song, including in original composition.

Cupeño is an agglutinative language in the sense summarized by Plank (1995), characterized by complex words consisting of long strings of affixes that largely retain a CV or CVC syllable structure, being loosely bound to word roots and to one another with relatively few morphologically conditioned word-internal alternations. (1) illustrates a minimal sample of the rich morphological apparatus for verb constructions (primary stress is on the first syllable of the word, unless it is marked with an underline; the symbol 'e' stands for a central vowel /ə~i/.)

- (1a) mi = pem-chi'
 them = they-gather
 'they gathered them'

- (1b) mi = ne-chi'-qal
 them = I-gather-IMPERF.SING
 'I was gathering, I would gather'
- (1c) mi = chem-chi'-lyu-wen
 them = we-gather-go.to-IMPERF.PL
 'we were going in order to gather them, we would go in order to gather them'

The examples in (2) illustrate nominal constructions. As in most Uto-Aztecan languages, nouns (except Spanish loans and a few words for plants and animals) must appear with one of the nonpossessed noun suffixes (in (2), the suffix *-ly(a)*). Animate nouns are always marked for object case, in which case demonstratives, quantifiers, and adjectives appearing with them in complex nominal constructions also bear object case (and plural number suffixation if the noun is plural). With inanimate nouns, only qualifying elements appear with the object case marker; the noun itself is usually unmarked.

- (2a) axwesh achi-ly
 that pet-NON.POSSESSED
 'that pet'
- (2b) axwēchi-m ash-ly-a-m
 that-PL pet-NON.POSSESSED-PL
 'those pets'
- (2c) axwesh-m-i ash-ly-a-m-i
 that-PL-OBJ pet-NON.POSSESSED-PL-OBJ
 'those pets (object case)'
- (2d) i'i ne-ash
 this my-pet
 'this pet of mine'
- (2e) ivi-y ne-ach-i
 this-OBJ my-pet-OBJ
 'this pet of mine (object case)'

(3) ne-chi-'a
my-gather-POSSESSED
'my harvest'

(4) ne-'ash-lyu = 'ep
I-pet-VERBALIZER = REALIS
'I had a pet'

(5) i'i i-iv-i'aw ku'a-l hiw-qa ne-qwa'i-i'aw
this this-at fly- stand- my-food-at
 POSSESSED PRES.SING
'this fly is here on this food of mine'

(6) ne'=ne mukikma-l-i
I=I.ERG bird-NON.POSSESSED-OBJ
pe-wek-'i-y tew-qa'
its-wing-POSSESSED-OBJ see-PRES.SING
'I see the bird's wing'

(7) hani=**qwe**=**n**=**pe**
 exhort=NONINSTANTIATIVE =I.ABS=IRR
 nangini met'ish
 PAY.HABILITATIVE MUCH
 me=**qwe**=**pe** ichaa
 AND=NONINSTANTIATIVE =IRR GOOD
 miyax-wene
 BE-CUSTOMARY.STATIVE
 'if I had paid more it would be better for me'

nominative) encoding the person and number of the subject. In tenses other than the past, null subjects are common, although subject number is encoded in the present tense and imperfective aspect suffixes. However, multiple marking of subject values can also occur, with independent nouns and pronouns, clitics in the auxiliary complex or past-tense subject markers, and subject-number-marking aspect suffixes cooccurring in a single sentence. All verbs can appear with object proclitics (seen in (1)) that encode the person and number of the object. In non-past-tense verbs, clitics in the second-position auxiliary complex encode the person, number, and case of the subject. With imperative verbs, these clitics encode the object. However, the language also has dependent marking for case, with a generalized object case suffix $-i \sim -y$ on quantifiers, demonstratives, adjectives, and nouns and pronouns, as seen in the examples in (2) and (5).

The many unusual typological properties of Cupeño hint that the language has undergone esoterogeny (Thurston, 1987), accumulating strategies for distinction from its neighboring languages. The split-ergative case system, the exuberant development of discontinuous constituency, and the Suffixaufnahme found in possessive expressions in questions are unattested elsewhere in Uto-Aztecan. Esoterogeny, using Thurston's characterization, is exactly what would be predicted in a language with very few speakers – probably never more than 1000 – incorporated into the linguistic ecology of aboriginal California, a classic example of an accretion or residual linguistic zone (Nichols, 1992). Golla (2000) has observed similar processes of accumulated distinctiveness in the California Athabascan languages, the other major case in which a language family that is widespread outside aboriginal California has a few members within that zone.

Barragan L (2003). 'Movement and allomorphy in the Cupeño verb construction.' In Haugen J & Barragan L (eds.) *Studies in Uto-Aztecan*. MIT working papers on endangered and less familiar languages no. 5. Cambridge, MA: MIT Department of Linguistics. 141–162.

- Bright W & Hill J H (1967). 'Linguistic history of the Cupeño.' In Hymes D H & Biddle W (eds.) *Studies in southwestern ethnolinguistics*. The Hague: Mouton. 352–391.
- Golla V (2000). 'Language history and communicative strategies in aboriginal California and Oregon.' In Miyaoka O & Oshima M (eds.) *Languages of the North Pacific rim*, vol. 5. Suita, Japan: Osaka Gakuin University, Faculty of Informatics. 43–64.
- Hill J H (2003). 'Subject number, verb agreement, grammaticalization and transitivity in the Cupeño verb construction.' In Carnie A, Harley H & Willie M (eds.) *Formal approaches to function in grammar: papers in honor of Eloise Jelinek*. Philadelphia: John Benjamins. 207–226.
- Hill J H (2005). *A grammar of Cupeño*. University of California publications in linguistics. Berkeley, CA: University of California Press.
- Hill J H & Nolasquez R (1973). *Mulu'wetam: the first people*. Banning, CA: Malki Museum Press.
- Hinton L (1991). 'Takic and Yuman: A study in phonological convergence.' *International Journal of American Linguistics* 57, 133–157.
- Hyer J (2001). 'We are not savages': *Native Americans in southern California and the Pala Reservation, 1840–1920*. East Lansing, MI: Michigan State University Press.
- Jacobs R (1975). *Syntactic change: a Cupan case study*. University of California publications in linguistics 79. 3–110 Berkeley, CA: University of California Press.
- Nichols J (1992). *Linguistic diversity in space and time*. Chicago, IL: University of Chicago Press.
- Plank F (1995). '(Re-)introducing *Suffixaufnahme*.' In Plank F (ed.) *Double case: agreement by Suffixaufnahme*. New York: Oxford University Press. 3–110.
- Thurston W F (1987). *Processes of change in the languages of northwestern New Britain*. Canberra, Australia: Pacific Linguistics.

Cushitic Languages

D Appleyard, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

The Cushitic languages are generally thought of as forming a distinct family of the Afroasiatic superfamily or phylum, comprising four branches in distribution from north to south: Beja, Central Cushitic or Agaw, East Cushitic, and South Cushitic. Of these, East Cushitic is by far the largest both in terms of number of languages and of the overall number of speakers of those languages. East Cushitic is also the most complex branch insofar as it is further divided into several discrete sub-branches: Saho-Afar, Lowland East Cushitic, Highland East Cushitic, and as has been suggested Dahalo, a single language formerly subsumed under South Cushitic. Indeed, some now prefer to see South Cushitic (minus Dahalo) as a further sub-branch of East Cushitic and not a separate branch of the family (see **Figure 1**).

In terms of numbers of speakers, most Cushitic languages are comparatively small, with a few thousands or tens of thousands of speakers, and occasionally with only a couple of hundred or fewer. Although available figures are not always reliable, the only Cushitic languages with more than 1 million speakers are Afar (1.5 million), Oromo (at least 18 million, all varieties), Sidamo (1.8 million), and Somali (between 10 and 11 million). The principal branches of the Cushitic family are as follows (see also **Figure 1**):

- Beja (or Bedawi, Bedawiye), though showing some dialect variation, is regarded as a single language and is the sole representative of North Cushitic.
- Central Cushitic or Agaw forms a fairly cohesive group of four small languages or dialect clusters, one of which, Bilen (Bilin), is spoken in Eritrea, the others in Ethiopia. The largest language is Awngi with about 300 000 speakers.
- The Highland East Cushitic group comprises five major languages with some variants that are sometimes considered separate languages, all spoken in the Rift Valley region of Ethiopia. The largest language is Sidamo (Sidaama) with about 1.8 million speakers, followed by Hadiyya with just under 1 million.
- The Lowland East Cushitic branch has the largest number of languages, about fifteen, stretching from Eritrea in the north to the south of Ethiopia, Somalia, and Djibouti, and beyond into Kenya and Tanzania. The Cushitic languages of Kenya are, in addition to Dahalo, extensions of those spoken to the north in Ethiopia and Somalia, and are all varieties of the two large Lowland East Cushitic languages, Oromo and Somali. The few Cushitic languages of Tanzania spoken by few people all belong to the South Cushitic branch, except for the probably extinct Yaaku, which has been seen as forming a discrete branch of Southern Lowland East Cushitic, perhaps linked to the Dullay (previously called Werizoid) languages of Ethiopia. Oromo and Somali are the languages with the

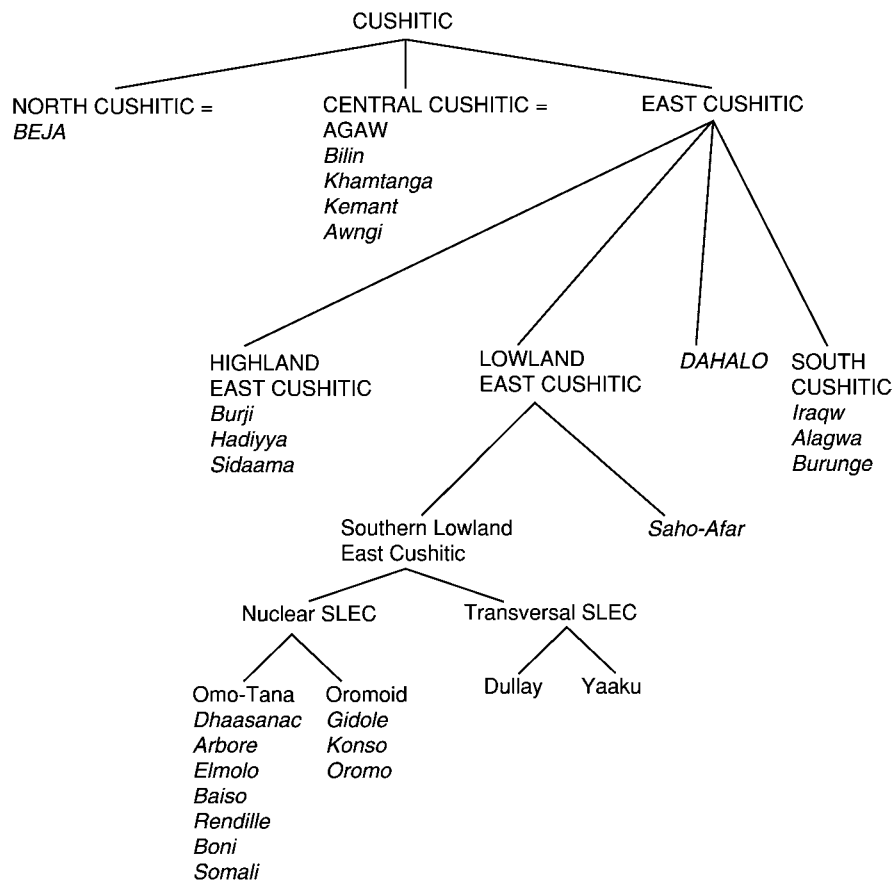


Figure 1 The Cushitic languages.

largest populations, at least 18 and 11 million speakers, respectively. Afar is the only other Lowland East Cushitic language with more than 1 million speakers, forming a separate sub-branch with the closely related Saho.

- The South Cushitic languages, all of which are spoken in Tanzania, and about whose position within the family there has been some debate, are all minority languages, and several are extinct or severely threatened.

The Question of Omotic

The ongoing re-analysis of the internal classification of Cushitic is not the only question regarding the nature of the family, nor the most recent one. For many years since the first attempts at classification of Cushitic, a further branch called West Cushitic was proposed, comprising a number of languages spoken in South West Ethiopia. There are sufficient substantial differences both in morphology and lexicon that set these languages apart from the rest of Cushitic, such that the erstwhile West Cushitic, now renamed Omotic, was proposed as a quite separate family of

the Afroasiatic phylum (definitively in Fleming, 1969), and the majority of linguists working in the area now concur with this classification. It has also been suggested that only part of Omotic, the Aroid (also called Ari-Banna, or Southern Omotic) languages, form a separate branch of Afroasiatic, while the rest are part of Cushitic. These problems of classification essentially revolve around the questions (a) of how much that is similar between Omotic and Cushitic is due to shared archaisms from Afroasiatic, and (b) how much arises from convergence due to an extended period of geographical proximity. There are certainly many similarities at all levels of linguistic analysis that are best explained by contact and convergence. Further discussion of Omotic is excluded from what follows.

Typological Characteristics of Cushitic Languages

While there is considerable variety in details of linguistic types among the Cushitic languages, it is by and large possible to draw up a list of structural features that exemplify most Cushitic languages. This

was done most clearly by Hetzron (1980: 14–53), and though some of the details are rather language-specific, what was said there is a sound statement. The features that he lists are all morphological, and it is indeed at the level of morphology (inflectional and derivational) that most of the strongest diagnostics can be found. A slightly different list is presented below, differing from Hetzron's principally only in its wider scope. Not all Cushitic languages obviously exhibit all of these features, but it is fair to say that a good part of them can be found or is traceable in probably all languages, and certainly in what Hetzron called the 'safe' branches of the Cushitic family: Agaw, Highland East Cushitic (his Rift Valley Cushitic), and Lowland East Cushitic (i.e., excluding South Cushitic). Lastly, what follows is by its very nature a generalization, and grammars of individual languages should be consulted for details.

Phonology

- a range of special coronal and velar consonants with secondary articulation, typically glottalized or implosive: e.g., t', d, ts', tʃ', k', g'...
- a tone-accent system with an underlying high-low contrast, functioning more in the context of morphological marking than lexically: e.g., Somali libaax = /libæːh/ 'lion' L.L (subject case), L.HL (non-subject case), inan = /inan/ H.L (masc. 'boy'), L.L (fem. 'girl').

Morphology

- two genders, masculine and feminine, the latter being the marked form, either by suffix and/or tone, with concord marking in pronouns and verbs, and with gender marking in pronominal (demonstrative) particles being widely masc. k-, fem. t-.
- a primary case system with two terms: a marked nominative or subject case (probably originally only on masc. nouns), and an unmarked absolutive case. A special genitive or possessive case may probably be also added to the primary cases. Other case functions are variously expressed by postpositions, evolving variously into suffixes or verbal proclitics. The morphemes expressing these various case functions are also by and large common between languages: e.g., a dative/benefactive -s, also -k in some languages; a locative/allative -l or -d/-t, sometimes appearing with comitative function. The marked subject case is often -i, though the vowel -u is also used in some languages.
- heterogeneous noun plural formation by means of a wide array of suffixes and occasionally a

degree of internal modification. In some languages, Lowland East Cushitic especially, gender polarity can also be observed between singular and plural. The quantificational system of nouns often also includes categories such as collective, singulative, and paucal: e.g., Hadiyya fellaʔa 'goat(s)', fella-kkitʃtʃo '(one) goat', fella-ʔuwwa '(lots of) goats', fella-kkitʃtʃaʔa '(a few) goats'.

- person marking on verbs for subject with seven terms, gender being marked only in the 3rd person singular. (Here Beja differs, marking gender also in the 2nd singular.) The patterning of person marking morphemes employed is perhaps the most immediately visible diagnostic of the Cushitic languages, and indeed of their membership in the Afroasiatic phylum (see **Figure 2**). In some languages (e.g., Beja, Afar, Somali), there are two systems of person marking, one by means of prefixes to the verb stem (combining with suffixes for marking plural in the 2nd and 3rd persons), and the other by means of suffixes placed after the verb stem. The former is the more archaic, having direct correspondents in Berber and Semitic, while the second is a specifically Cushitic innovation, whereby an auxiliary inflecting after the archaic pattern became fused with the verb stem as a suffix element carrying both markers of person and tense-mood-aspect. The actual markers of person are the same in both types, except for the 3rd masc. and pl. where prefix y->Ø. **Figure 3** shows prefix and suffix inflecting paradigms in two tenses in Afar; the suffix -[V]h is required here in final position where the verb is focused.
- the finite verb has three primary tense-mood-aspect (TMA) forms: a past or perfective, a non-past or imperfective, and a subjunctive which typically has both modal and dependent functions. These are distinguished by vocalic variation, either in the suffixes in suffix-inflecting verbs, or originally by internal vocalic modification of the verb stem in prefix-inflecting verbs. In addition to these three primary forms, most languages have developed a range of other TMA forms, often including distinct negative paradigms, sometimes employing

| | SING. | PLUR. |
|--------|-------|--------------|
| 1 | ʔ | n |
| 2 | t | t ... -n |
| 3MASC. | y > Ø | y > Ø ... -n |
| 3FEM. | t | |

Figure 2 The Cushitic 'block pattern' of person marking.

| | prefix perfect | prefix imperfect | suffix perfect | suffix imperfect |
|--------|----------------|------------------|----------------|------------------|
| 1sg. | u-duure-h | a-duure-h | fak-e-h | fak-a-h |
| 2sg. | tu-duure-h | ta-duure-h | fak-te-h | fak-ta-h |
| 3m.sg. | yu-duure-h | ya-duure-h | fak-e-h | fak-a-h |
| 3f.sg. | tu-duure-h | ta-duure-h | fak-te-h | fak-ta-h |
| 1p1. | nu-duure-h | na-duure-h | fak-ne-h | fak-na-h |
| 2p1. | tu-duuree-n-ih | ta-duuree-n-ih | fak-teen-ih | fak-taan-ah |
| 3pl. | yu-duuree-n-ih | ya-duuree-n-ih | fak-een-ih | fak-aan-ah |

Figure 3 Person marking in the verb in Afar.

auxiliaries and sometimes additional suffix elements. So, for example, alongside Afar fakeh and fakah, etc., in **Figure 3**, there are also such forms as jussive fakay ‘let me open’, requestive fako ‘may I open?’, negative perfective ma-fakinniyo ‘I didn’t open’, anticipatory fake-liyo ‘I will open’, negative present continuous fakah-maan ‘I am not opening’, and so on.

- the verb has a rich system of stem derivation expressing voice with such categories as passive, causative, autobenefactive or middle, reciprocal, etc. The markers of voice, which generally follow the lexical stem of the verb and come before person and TMA markers in suffix-inflecting verbs, show a marked degree of commonality between all Cushitic languages, with the causative marked by –[V]s, the passive by –[V]m, and the autobenefactive or reflexive by –[V]t, or elements that can be shown to have derived from such. Combinations of derivational suffixes also occur. Stem reduplication or partial reduplication is also typically employed in making iteratives, or sometimes reciprocal forms, often in combination with the primary stem derivational suffixes.

Syntax

- focus systems are common, often deriving from cleft constructions; special reduced subject-focus verb paradigms are also common: e.g., Oromo

obboleettii issa-tu foon nyaat-a
sister.ABS his-FOCUS meat eat-3MASC.SING.IMPERF
 ‘it is his sister who eats meat / **his sister** eats meat’ i.e.,
 verb 3masc.sing and not in agreement with focused
 subject.

Contrast the same sentence with predicate focus:

obboleettii-n isaa foon ni nyaat-ti
sister-SUBJ his meat FOCUS eat-3FEM.SING.IMPERF
 ‘his sister **eats meat**’

- clause chaining constructions, often including the use of special converbs, e.g., Hadiyya:

it’ti-m sigg-aa wo?oo-ma
it-and cool-3MASC.SING.CONVERB water-DEF
 t’uut’-aa lasage k’ure?e
suck-3MASC.SING.CONVERB after pot

giira-nne kaas-akkamo
fire-LOC put-3POL.IMPERF

‘and after it has cooled and absorbed the water, one
 puts the pot on the fire’

Additionally, subordinate clauses often use special verb forms that are either of a relative clause verb type, or are ostensibly derived from relative constructions.

- sentence word order is SOV, though both head-final and head-initial types of phrase (e.g., noun phrases) occur.

Bibliography

- Appleyard D L (1986). ‘Agew, Cushitic and Afroasiatic. The personal pronoun revisited.’ *Journal of Semitic Studies* 31, 195–236.
- Appleyard D L (1996). ‘The position of Agaw within Cushitic.’ In Zemánek P (ed.) *Studies in Near Eastern languages and literatures: Memorial volume of Karel Petrůček*. Prague: Academy of Sciences of the Czech Republic.
- Bender M L (ed.) (1976). *The Non-Semitic languages of Ethiopia*. East Lansing: African Studies Center, Michigan State University.
- Fleming H (1969). ‘The classification of West Cushitic within Hamito-Semitic.’ In McCall D F, Bennett N R & Butler J (eds.) *Eastern African history*. New York: Frederick Praeger. 3–27.
- Hetzron R (1980). ‘The limits of Cushitic.’ *Sprache und Geschichte in Afrika* 2, 7–126.
- Sasse H-J (1980). ‘Ostkuschitische und semitische Verbalklassen.’ In Diem W & Wild S (eds.) *Studien aus Arabistik und Semitistik*. Wiesbaden: Harrassowitz. 153–174.

- Sasse, Hans-Jürgen (1981). 'Die kuschitischen Sprachen.' In Heine B, Schadeberg T C & Wolff E (eds.) *Die Sprachen Afrikas*. Hamburg: Helmut Buske. 187–212.
- Tosco M (2000). 'Cushitic overview.' *Journal of Ethiopian studies* 33, 87–121.
- Voigt R (1996). 'Zur Gliederung des Kuschitischen: Die Präfixkonjugationen.' In Griefenow-Mewis C & Voigt R M (eds.) *Cushitic and Omotic languages. Proceedings of the Third International Symposium (Berlin, March 17–19, 1994)*. Köln: Rüdiger Köppe. 101–131.
- Zaborski A (1975). *The verb in Cushitic*. Kraków: Zeszyty Naukowe Uniwersytetu Jagiellońskiego, Prace Językownawcze.
- Zaborski A (1984). 'Remarks on the genetic classification and the relative chronology of the Cushitic languages.' In Bynon J (ed.) *Current progress in Afro-Asiatic linguistics*. Amsterdam: John Benjamins. 121–138.

Czech

L A Janda, University of North Carolina, Chapel Hill, NC, USA

© 2006 Elsevier Ltd. All rights reserved.

Czech is the official language of the Czech Republic (with over 10.2 million inhabitants), bordered by Austria, Germany, Poland, and Slovakia. There are significant emigré populations, particularly in the United States, Canada, and Australia.

Czech is a West Slavic language (with Slovak, Sorbian, and Polish). In 862 the ancestors of the Czechs became the first Slavs to achieve literacy in their own language when the Byzantine Saints Cyril and Methodius brought liturgical texts translated into Old Church Slav(on)ic. In the 12th to 14th centuries Czech underwent the 'přehláska' vowel-fronting changes that established 'hard' vs. 'soft' stem differentiations throughout the morphology. The 15th-century theologian Jan Hus is credited with the invention of diacritic marks to adapt the Latin alphabet to Czech phonology. Under the control of the Habsburg dynasty, particularly after the 1620 defeat at White Mountain, use of German was enforced at the expense of Czech. Czech endured decline and disuse before reasserting itself as a literary and official language in the early 19th century. The modern literary language is based on the 16th-century *Kralice* Bible, but vernacular Czech had continued to evolve, resulting in a pronounced gap between the literary and spoken language (involving phonology, morphology, syntax, and lexicon), often described as diglossia between Literary Czech (LCz) and Colloquial Czech (CCz).

Most peripheral zones of the Czech Republic belong to no dialect group due to resettlement by Czech speakers from other locations after German inhabitants were ousted at the end of World War II. The two largest dialect groups are classified according to their treatment of certain etymologically

long vowels as Bohemian (central and western) and Hanák Moravian (eastern). Northeastern Lachian Silesian and mixed Polish-Czech dialects serve as a transition to Polish, characterized by loss of vowel length, penultimate stress, and consonantism similar to Polish. Southeastern Moravian-Slovak dialects serve as a transition to Slovak, characterized by retention of *û/ú* and of back vowels after palatal consonants.

Czech has the following consonant phonemes: voiced and voiceless bilabial, dental, palatal, and velar plosives; bilabial, dental, and palatal nasals; a dental trill; voiced and voiceless labiodental, dental, and postalveolar fricatives; a voiceless velar fricative; a voiced glottal fricative; voiceless dental and palatal affricates; a palatal approximant; and a dental lateral approximant. In addition, Czech has a double-articulation phoneme produced by simultaneous pronunciation of the dental trill and the voiced postalveolar fricative (*ř*). Final devoicing (*dub* [dup] 'oak') and regressive voicing assimilation of obstruents (*kdo* [gdo] 'who') is pervasive, and progressive devoicing occurs in certain word-initial clusters. Two subphonemic consonants are the velar nasal (an allophone of *n* before a velar, as in *banka* 'bank') and the glottal plosive, which appears before word-initial vowels and between vowels at the prefix boundary (eliminating vowel chains in Czech). Czech has a five-vowel system, consisting of short *a, e, i/y, o, u* and long *á, é, í/ý, ó, û/ú*. There are seven native (*ij, ej, aj, oj, uj, új, ou*) and two borrowed (*eu, au*) diphthongs. The liquids *r* and *l* participate in both syllable peaks (as vowels in *smrt* 'death' and *vlk* 'wolf') and slopes (as consonants). The sole phonemic prosodic feature is vowel length. A non-phonemic stress falls on the first independent syllable of a phonological word (which may contain stressless proclitics and enclitics). CCz shows reflexes of *ě>í/ý* and *ý>ej*. Since etymological *é* and *ý* figure as essential components of Czech morphology, these vowel

changes are prominent in differentiating CCz and LCz morphology. Another hallmark of CCz is prothetic *v*-before word-initial *o*-, as in CCz *vocas* 'tail' (cf. LCz *ocas*).

Inflectional morphology is expressed as synthetic terminal desinences added to the stems of nouns, adjectives, verbs, and most pronouns. Inflectional desinences conflate all relevant categories: gender, number, and case for nouns and adjectives; person and number for non-past conjugations; and gender, person, and number for past conjugations. All native stems are inflected, as are the vast majority of foreign borrowings. Morphophonemic alternations include: vowel-zero alternations (*pēs* 'dog'Nsg : *psi* 'dogs'Npl); qualitative vowel alternations (*moucha* 'fly'Nsg vs. *práce* 'work'Nsg); quantitative vowel alternations (*nést* 'carry' vs. *nesu* 'I carry'); and consonant alternations (*kniha* 'book'Nsg vs. *knize* 'book'Lsg).

All nouns have grammatical gender (masculine, feminine, neuter), and are declined for both number (singular, plural) and case (nominative, genitive, dative, accusative, vocative, locative, instrumental). Each gender has its own set of characteristic paradigms, including hard-stem types, soft-stem types, and special types. Masculine paradigms regularly signal animacy with distinctive animate endings in the Dsg, Asg, Lsg, and Npl. There are also special paradigm types that signal virile (male human) gender.

Adjectives are declined to match the gender, case, and number of the nouns they modify. Like nouns, adjectives have both hard- and soft-stem paradigms.

Pronouns have a mixed declensional type, using endings from both noun and adjective paradigms. Personal pronouns have both full (emphatic) and enclitic short forms.

Cardinal numerals are inflected for case, *jeden* 'one' and *dva* 'two' additionally distinguish gender,

and *jeden* 'one' distinguishes number as well. Ordinal numerals are declined as adjectives.

Verbal morphology expresses aspect (perfective, imperfective; obligatory for all forms), mood (indicative, imperative, conditional), voice (active, passive), tense (non-past, past), person (first, second, third), gender (masculine animate and inanimate, feminine, neuter), and number (singular, plural); motion verbs distinguish directionality. Past conjugation uses the auxiliary verb *být* 'be' in the first and second persons. As a rule, non-past conjugation of perfective verbs signals future tense, whereas non-past conjugation of imperfective verbs signals present tense. Imperfective verbs form a periphrastic future tense with forms of *být* 'be'. Most simplex verbs are imperfective (*volat* 'call'), but some are perfective (*dát* 'give'). Perfective and imperfective verbs can be derived from simplex verbs by means of prefixation and suffixation.

Czech is a pro-drop language; nominative case pronouns are emphatic. Czech case indicates the syntactic function of a given noun phrase and the relationship it bears to the verb and to other noun phrases and can also indicate pragmatic relationships. Word order is free, however there is a strict order of enclitics after the first stressed word in a clause.

Bibliography

- Havránek B & Jedlička A (1981). *Stručná mluvnice česká*. Prague: Státní pedagogické nakladatelství.
- Janda L A & Townsend C E (2000). *Czech*. (= *Languages of the World/Materials* 125). Munich: LINCOM EUROPA.
- Karlík P *et al.* (1995). *Příruční mluvnice češtiny*. Prague: Lidové noviny.
- Palková Z (1997). *Fonetika a fonologie češtiny*. Prague: Univerzita Karlova.
- Townsend C E (1990). *A description of spoken Prague Czech*. Columbus, OH: Slavica.

This page intentionally left blank

D

Danish

T Lundskær-Nielsen, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Danish is the native language of more than 5 million people in the Kingdom of Denmark, including the Faroe Islands and Greenland, where Danish is the second language for most of the inhabitants. It is the first language of some 50 000 people in South Schleswig, North Germany, and of more than 100 000 Danes currently living in other European countries, including Norway and Sweden. There are also some Danish émigré communities in the United States and Canada.

History

Danish belongs to the North Germanic group of the Germanic languages. The earliest language of this group, Ancient Scandinavian (c. 200–600), was common to the Scandinavian area, as indicated by runic inscriptions. Around the 9th century, the descendant language, Old Scandinavian, gradually developed into two distinct branches, viz., West Scandinavian (Old Norwegian and Old Icelandic (or Old Norse)) and East Scandinavian (Old Danish and Old Swedish). Historically, the Danish language may be divided into three periods: Old Danish (c. 800–1100) spans the Viking Age, Middle Danish (c. 1100–1500) was the language of the late Middle Ages, and Modern Danish covers the time from around the Reformation (and the first translation of the Bible) to the present. Of the Scandinavian languages, Danish is now closest to Norwegian *bokmål* and Swedish. For more than 400 years (1380–1814), Norway was part of a Dual Monarchy under the Danish Crown.

Early changes from Old Scandinavian to East Scandinavian saw some diphthongs develop into monophthongs (e.g., <ai> and <ei> > <e> and <au> and <øy> > <ø>) and the loss of <h> before <l, n, r>. In Middle Danish, a number of changes in sounds and spelling began to distance Danish (Da.) from Swedish (Sw.): (1) the full vowels <a, i, u> were weakened

to <e> [ə] in unstressed syllables (e.g., Sw. *skriva*, Da. *skrive* ‘write’); (2) the aspirated stops <p, t, k> became unaspirated <b, d [ð], g [j]> (e.g., Sw. *gata*, Da. *gade* ‘street’); and (3) the sounds [d, g, v], when they did not disappear altogether, changed into [ð, j, w], respectively, with [j] and [w] often becoming the second element of diphthongs (e.g., Da. *dag* [daj] ‘day,’ *liv* [liw] ‘life’).

Orthography

Modern Danish has the same 26 letters as English has, plus the three additional vowels, <æ> [ɛ], <ø> [ø], and <å> [ɔ], which are placed last in the alphabet; the letters <q, w, x, z> occur only in foreign loans. Since there have been very few and only minor spelling reforms for centuries, Danish spelling does not accurately reflect the pronunciation. This is true concerning both consonants and vowels.

Pronunciation

Danish has some 15–20 consonant phonemes, comprising at least the stops /p t k b d g/, the fricatives /f v s ð/, the nasals /m n ŋ/, the lateral /l/, the uvular /r/, the glottal /h/, and the two ‘semivowels’ /j/ and /w/. All the stops are voiceless, so /p t k/ and /b d g/ are distinguished solely by /p t k/ having (strong) aspiration and by /b d g/ being unaspirated. However, in positions other than initially before /j, l, r, v/ and/or a full vowel, /p t k/ are pronounced [b], [d] or [ð], and [g]. Postvocalic /r/ becomes vocalic [ʀ], merging with the preceding vowels, e.g., in the <-(e)r> ending of the present tense and the plural of some nouns, as in *jeg læse-r* /'lɛ:ʂ/ ‘I read’ (cf. *at læse* ‘to read’) and *mange sted-er* /'sde:ð/ ‘many places’ (cf. *et sted* ‘a place’). The /h/ is pronounced initially only before a full vowel and is dropped before /j/ or /v/, as in *hjelpe* /jɛlb/ ‘home’ and *hvad* /vɑð/ ‘what.’

Danish has 11 vowel phonemes /i e ɛ ɔ y ø ʊ o ɔ/, all of which have a long and a short realization, so the real number may be said to be 22. There are eight front vowels – five unrounded /i e ɛ ɔ/ and three rounded /y ø ʊ/ – and three rounded back

vowels (/u o ɔ/). There are no unrounded back vowels. There are also some allophones, e.g., [œ] in relation to [æ], and [ɐ] in relation to [ɔ], both lowered by an adjacent /r/. The unstressed vowels [ə] and [ʌ] may be seen as allophones of /e/ and /r/, respectively. The number of front vowels (unrounded and rounded) is very large compared with most other European languages. In addition, there are two sets of diphthongs with an underlying long or short vowel, respectively, as the first element, and /j/, /w/, or /ʌ/ as the second, numbering over 20 in all.

Danish has a unique feature called *stød* (marked /ʔ/), which resembles the glottal stop in English but is more of a 'creaky voice' without complete closure of the vocal cords. It can occur when there is a so-called *stød* base in the form of either a long vowel or a short vowel + a sonorant (/l/ or a nasal), as in *hus* /hu:ʔs/ 'house' and *lyn* /lynʔ/ 'lightning.' Certain word pairs are distinguished in speech solely by the presence or absence of *stød* (*hund* /hunʔ/ 'dog' vs. *hun* /hun/ 'she'). Some southern Danish dialects are without *stød*, though.

There are no tones and no sentence accent in Danish, so the last stressed syllable shows no more prominence than other stressed syllables do. This can make Danish speakers sound rather dull and uninteresting to foreigners. The intonation contour of the stressed syllables is characterized by a gradual fall, but the first unstressed syllable in a prosodic stress group is on a higher pitch than the immediately preceding stressed one is. Both range of and variation in pitch are much narrower than in English, Norwegian, and Swedish.

Morphology

Danish nouns are inflected for number, gender, and case. There are two numbers (singular and plural), two genders (common and neuter), and two cases (unmarked case and genitive). Plural endings are *-e*, *-(e)r*, or *zero*-ending, though some foreign loans may retain a foreign ending, as in *faktum*, *fakta* 'fact(s)' and *fan*, *fans* 'fan(s).' The indefinite article is *en* in common gender and *et* in neuter (*en bil* 'a car,' *et dyr* 'an animal'); the definite article is either a front article (*den* or *det* (SG), *de* (PL)), used when an adjective follows the article, as in *den store bil* 'the big car,' *det store dyr*, *de store biler/dyr* 'the big car(s)/animal(s),' or an end article attached to the noun (*-(e)n* or *-(e)t* (SG), *-ne* (PL)), when there is no adjective, as in *bil-en* 'the car,' *dyr-et* 'the animal,' *biler-ne/dyre-ne* 'the cars/animals.' The genitive ending is *-s* (*bilen-s lygter* 'the car's lights').

Verbs have no person or number distinction and thus no agreement with the subject, as in *jeg/hun/de er/spiser* 'I am/eat, she is/eats, they are/eat.' There are

four conjugations: three weak ones with the past tense endings *-(e)de*, *-te*, *-de* and one strong one with the *zero*-ending (*-t*), as in *leg-ed*, *hør-te*, *sag-de*; and *sang*, *fand-t* 'played, heard, said'; and 'sang, found.' The past participle ending is *-(e)t*, as in *leg-et* 'played and *hør-t* 'heard.' The infinitive ends in *-e* or a full vowel (*at leg-e* 'to play,' *at få* 'to get'), and the present participle ends in *-ende* (*løb-ende* 'running'). There are two passive forms, an *-s* passive and a form with the auxiliary verb *blive* + a past participle, as in *brevet sendte-s/blev sendt* 'the letter was sent.'

Most adjectives agree with nouns and articles and have the endings *zero* or *-t* (SG) or *-e* (PL) in indefinite forms, and *-e* in all definite forms: *god smag* 'good taste,' *god-t arbejde* 'good work,' *god-e kager* 'good cakes' (definite forms were mentioned previously). Comparison of adjectives is marked by the endings *-(e)r* (comparative) and *-(e)st* (superlative) or *-* with longer adjectives – the adverbs *mere* and *mest*. Most adverbs have the ending *-t* (*han løb hurtigt-t* 'he ran fast').

Personal pronouns show case distinction (nominative vs. oblique) as well as person and number distinction, as in *jeg/mig*, *hun/hende*, *de/dem* 'I/me, she/her, they/them.' Some possessive pronouns inflect like adjectives (*min*, *din*, *sin* 'my, your, his/her/its [third-person reflexive]') and others have just one form in all uses (*hans*, *deres* 'his, their').

Syntax

Danish word order is relatively fixed, but a distinction must be made between main clauses and subordinate clauses. A sentence schema, devised by the Danish linguist Paul Diderichsen, can account for the order of most Danish sentences. As shown in Table 1, the two types of clauses consist of seven positions (to which can be added extra positions both initially and finally), but note the different order of *v*, *n*, and *a* (finite verb, subject (when not in front), and central adverbial, respectively). In main clauses, another element may move to the front (F) position for emphasis (i.e., topicalization), thus causing the subject (here: *han*) to move into the *n*-position. Note that the finite verb is always in second position, because Danish is a V2 language (V is the nonfinite verb). Examples of A (other adverbials) (especially) or N (object/complement, both indirect and direct) moving to the front position are common:

| F | v | n | a | V | N | A |
|------------|-----|-----|-------|-------|------------------------|----------|
| Til jul | har | han | altid | sendt | sin søster et brev. | — |
| Sin søster | har | han | altid | sendt | et brev | til jul. |
| Et brev | | har | han | altid | sendt sin søster | til |

Table 1 Positions in main and subordinate clauses in Danish^a

| Clause | Position | | | | | | |
|--------------------------|-------------|------------|-----------------|-----------------|---------------|---|-----------------------------|
| Main | F | v | n | a | V | N | A |
| | Han (He | har has | — | altid always | sendt sent | sin søster (IO) et brev (DO) his sister a letter | til jul. for Christmas.) |
| Subordinate ^b | k | n | a | v | V | N | A |
| | at (that | han he | altid always | har has | sendt sent | sin søster (IO) et brev (DO) his sister a letter | til jul. for Christmas) |

^aAbbreviations: F, front position (subordinate clause: k, = conjunction); v, finite verb; n, subject (when not in F); a, central adverbial(s); V, nonfinite verb(s); N, object/complement; A, other adverbial(s). Note that an indirect object (IO) precedes a direct object (DO).

^bAssume a preceding main clause: *Han siger* 'He says' (note that there is no change of order in English!).

Examples with **a** or **VNA** (moving together) in **F** can also be construed, but are rarer.

Questions are formed either by inversion of subject (S) and finite verb (v) (Sv > vS), thus leaving **F** empty, or by having a question-word in **F** (e.g., *Hvorfor* 'Why'):

| | | | | | | |
|----------|----------|----------|----------|----------|------------|----------|
| F | v | n | a | V | N | A |
| — | Har | han | altid | sendt | sin søster | til jul? |
| | | | | | et brev | |
| Hvorfor | har | han | altid | sendt | sin søster | til |
| | | | | | et brev | |

When a main clause (MC) follows a subordinate clause (SC), there is inversion in the main clause:

| | | | | | | | |
|--|-----|-------|-------|-------------|------|-----|-------|
| | SC | | | | MC | | |
| Da | han | havde | sendt | brevet, | gik | han | hjem. |
| k | S | v | V | DO | v | S | A |
| (When | he | had | sent | letter-the, | went | he | home) |
| 'When he had sent the letter, he went home.' | | | | | | | |

Language Authorities

Dansk Sprognævn (the Danish Language Council), which acquired legal status in 1997, monitors the development of Danish, including the adoption of new loanwords. The Council provides guidance on language matters and is the highest authority on modern Danish spelling.

Bibliography

- Allan R & Lunds-kær-Nielsen T (2001). 'Danish.' In Garry J & Rubino C (eds.) *Facts about the world's languages*. New York/Dublin: New England Publishing Associates. 184–189.
- Allan R, Holmes P & Lunds-kær-Nielsen T (1995). *Danish. A comprehensive grammar*. London/New York: Routledge.
- Allan R, Holmes P & Lunds-kær-Nielsen T (2000). *Danish. An essential grammar*. London/New York: Routledge.

- Andersen S T (1998). *Talemåder i dansk. Ordbog over idiommer*. Copenhagen: Munksgaard.
- Becker-Christensen C (2004). *Politikens Nudansk syntaks*. Copenhagen: Politiken.
- Becker-Christensen C & Widell P (2003). *Politikens Nudansk grammatik* (4th edn.). Copenhagen: Politiken.
- Diderichsen P (1962). *Elementær dansk grammatik*. Copenhagen: Gyldendal.
- Elsworth B (1994). *Teach yourself Danish*. London: Hodder & Stoughton.
- Galberg Jacobsen H & Skyum-Nielsen P (1996). *Dansk sprog. En grundbog*. Copenhagen: Schønberg.
- Galberg Jacobsen H & Stray Jørgensen P (2002). *Håndbog i nudansk* (4th edn.). Copenhagen: Politiken.
- Gregersen F, Kristiansen T & Pedersen I L (1996). *Dansk sproglære*. Copenhagen: Dansk lærerforening.
- Grønnum N (2003). *Fonetik og fonologi* (3rd edn.). Copenhagen: Akademisk Forlag.
- Haberland H (1994). 'Danish.' In König E & van der Auwera J (eds.) *The Germanic languages*. London/New York: Routledge. 313–348.
- Hansen A (1967). *Moderne Dansk I–III*. Copenhagen: Grafisk Forlag.
- Hansen E (1997). *Dæmonernes port* (4th edn.). Copenhagen: Hans Reitzel.
- Hansen E & Lund J (1994). *Kulturens gesandter. Fremmedordene i Dansk*. Copenhagen: Munksgaard.
- Haugen E (1976). *The Scandinavian languages*. London: Faber & Faber.
- Jarvad P (1995). *Nye ord – hvorfor og hvordan?* Copenhagen: Gyldendal.
- Jarvad P (1999). *Nye ord. Ordbog over nye ord i Dansk 1955–1998*. Copenhagen: Gyldendal.
- Jones W G & Gade K (1993). *Colloquial Danish*. London/New York: Routledge.
- Karker A (1993). *Dansk i Tusind År. Modersmål-selskabet*. Copenhagen: C. A. Reitzel.
- Lund J (2001). *Sproglig status*. Copenhagen: Hans Reitzel.
- Politikens Etymologisk Ordbog* (2000). Copenhagen: Politiken. [Contains a survey of Danish language history by A. Karker.]
- Politikens Nudansk Ordbog* (2001). Copenhagen: Politiken.
- Preisler B (1999). *Danskerne og det engelske sprog*. Roskilde: Roskilde U. P.

Rasmussen J (2002). *Dansk fonetik i teori og praksis*. Copenhagen: Special-Pædagogisk Forlag.
Retskriveringsordbogen (2001). Copenhagen: Alinea/Aschehoug.

Skautrup P (1944–1970). *Det Danske sprogs historie* (vols. I–IV). Copenhagen: Nordisk Forlag.
 Sørensen K (1995). *Engelsk i Dansk – er det et must?* Copenhagen: Munksgaard.

Dardic

S Munshi, University of Texas at Austin, Austin, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

‘Dardic’ languages are spoken in northwestern Pakistan and Jammu & Kashmir state in India, and extend into Afghanistan. The region, commonly known as ‘Dardistān,’ i.e., ‘the land of the Dard (people),’ is composed of the whole mountainous territory of the Hindukush, Swāt, and Indus Kohistān, the valleys of the Karakoram, and the western Himalayas. Dardistān also includes some areas occupied by non-Dardic language speaking people. Situated between South and Central Asia, with Iranian languages on one side and Indo-Aryan on the other, Dardic languages are in contact with and influenced by languages of other language families, such as Sino-Tibetan, as well as the language isolate Burushaski. One of the characteristic features of Dardic languages is that they have similarities with both Indo-Aryan as well as Iranian, the two major branches of the Indo-Iranian languages.

Dardic languages were previously divided into three sub-groups: Kāfiri/Kāfir or (present-day) Nūristānī group, Khowār group, and the Dard group proper (Grierson, 1919; Kachru, 1969). However, scholars now believe that Nūristānī is a separate sub-group of Indo-Iranian, while other languages currently classified as Dardic are of Indo-Aryan origin (Morgenstierne, 1965; Bashir, 2003: 822). Based on historical sub-grouping approximations and geographical distribution, Bashir (2003) provides six sub-groups of the Dardic languages:

1. Pashai group, also called Laghmānī, Deganó, or Dehgānī (Chugani and Chalās-KuRangal forming the eastern dialects; Sum, Damench, and Upper and Lower Darra-i-Nur constituting the south-eastern dialects; and several western dialects)
2. Kunar group (Gawarbati, Shumashti, and Grangali-Ningalami classified as the Gawarbati-type; and Dameli)
3. Chitral group (Khowar and the Kalasha sub-group)

4. Kohistānī group (Tirahi; the Dir-Swat sub-group; and the Indus-Kohistānī sub-group)
5. Shina group (the Kohistān sub-group, including Chilāsī and other languages; the Astor sub-group including Astori, Drāsī, and other languages; Gilgit sub-group, including Gilgiti and Brokskat in addition to others; and Palula)
6. Kashmiri (Standard Kashmiri, Kashtwāri/Kishtwāri, Poguli, Sirāji, Rāmbani, and Bunjwali dialects).

Of these, only Kashmiri has a well-developed tradition of written literature dating back to the 13th century. Originally written in Shārada, the current officially recognized script for Kashmiri is a modification of Perso-Arabic/Nastālīq. Shina and Khowar have also developed a writing system by modifying the Perso-Arabic script.

Available information on the numbers of speakers of most Dardic languages is based on estimated figures. The total number of speakers is about 5000 for Grangali (in 1994; spoken in the valleys south of Pech River in Kandai, Afghanistan; *Ethnologue*, 2003); 5000–6000 or less for Kalasha (spoken in southern Chitral District in Pakistan, closely related with Khowar) and Dameli (spoken in Damel valley towards the left bank of Chitral River); 8000–10 000 for Gawarbati (mainly spoken in Afghanistan; some speakers were also displaced to Pakistan during war); 60 000 for Torwali (Rensch, 1992: 33; Bashir, 2003: 864; spoken in Swat valley and Chail side valley; most speakers are bilingual in Pashto, and more and more are becoming bilingual in Urdu); 60 000–70 000 for Swat-Dir Kohistānī (in 1995; Baart, 1997: 4; spoken in Swat Kohistan and Dir Kohistan; most speakers are bilingual in Pashto); 200 000 for Indus-Kohistānī (Hallberg, 1992: 89; spoken in District Kohistan); 300 000 for Khowar (spoken in Chitral; some speakers are also found in Yasin and Ishkoman, upper Swat, Peshawar, and Karachi); and over 4 million for Kashmiri (*Ethnologue*, 2003; Koul, 2003: 897; spoken in India, primarily in Kashmir valley and its surroundings, and also in Pakistan-administered Kashmir; most speakers are bilingual in Urdu and sometimes Punjabi or other languages).

Estimates for total Shina speakers in Pakistan vary greatly, from 0.5 million (Radloff, 1992: 93) to about 3 million (Schmidt, 1988: 107–108). There are approximately 20 000 Shins (Shina speakers) in India (Radloff, 1992: 93). Shina is spoken in Gilgit, Hunza, Astor valley, Tangir-Darel valley, Chilas, Indus-Kohistān, and in the gorges of Brog-yul in central Ladakh south of the Hindukush-Karakoram ranges. The inhabitants of Brog-yul, speaking the Brokskat dialect of Shina, prefer to be called Shins/Shrins but they are popularly known as Brokpas/Dokpas by their Ladakhi and Balti neighbors (Sharma, 1998: 1). Most Shina speakers are bilingual in Balti, Kashmiri (eastern dialects), Burushaski and Khowar (Gilgit dialect), and Pashto and Indus-Kohistānī (Kohistānī dialects) (Bashir, 2003: 878).

History and Development

‘Dardic’ is a cover term used for a group of geographically contiguous languages of Indo-Iranian origin that share several linguistic features characteristic of themselves. It is derived from another term ‘Dard’ (dental *d*), which was originally used to refer to an ancient tribe living in the present-day Dardistān. Dards have been variously mentioned in literature (Ptolemy’s ‘Daradrai,’ Strabo’s ‘Derdai,’ the ‘Dardæ’ of Pliny and Nonnus, and Dinsyios Periêgêtês’ ‘Dardanoi’; Grierson, 1919: 1). ‘Dārada’/‘Darada’ have also been referred to in Sanskrit literature (e.g., ‘Dārada’/‘Darada’ by Kalhana in *Rājatarangini*). In Sanskrit the term *Dard* means ‘mountain’ and was perhaps used because most of the Dardic area is mountainous (Kachru, 1969: 285). Mohi-ud-Din (1998: 19) points to the possibility that the term *Dard* may be a corruption of *Dravād*, given the historical evidence that Dravidians inhabited a vast area, including northern India, before the advent of ‘Aryans’ into this land. He further claims that Dards were not an ‘Aryan’ race but they were the original inhabitants of this area while Aryans came later. The term *Piśachas* or *Paiśachas* (‘flesh devourers’), a derogatory term, also used in literature for Dards, was probably used by ‘Aryans’ to refer to the natives who perhaps called themselves *Dards*.

There has been a considerable debate over the classification of Dardic languages in terms of whether they are a third branch of Indo-Iranian language family (other two being Indo-Aryan and Iranian), or (at least, some of them) are of pure Indo-Aryan origin. Dardic languages have preserved many archaic Indo-Iranian features otherwise lost in the modern Indo-Aryan languages. A defining feature of Dardic languages is that they have undergone only some of

the major Middle Indo-Aryan (MIA) phonological and morphological changes. They have also developed certain areal features neither found in other Indo-Aryan (IA) nor in Iranian languages.

Phonological Characteristics

Dardic languages have descended from the north-western group of the MIA languages. Non-Dardic members of the same group include Punjabi, Sindhi, and Lahnda. One of the characteristic features of the phonological system of Dardic languages is the retention of the three-way distinction of Old Indo-Aryan (OIA) fricatives/sibilants -ś (palatal), s (dental), and ṣ (retroflex), which merged into one (dental s) or sometimes two (palatal ś and dental s) in many New Indo-Aryan (NIA) languages. For example, Pashai, Shumashti, Dameli, Khowar, Kalasha, Swat-Kohistani, Torwali, Indus Kohistani, and Shina have retained all three sibilants, while Grangali, Tirahi, and Kashmiri possess two sibilants (ś and s). Dardic languages have also retained the consonantal component *r* in the derivatives of the OIA syllabic *ṛ* that had a number of reflexes in MIA, viz., *a*, *i*, or *u*. Various OIA consonant clusters lost in other IA languages are retained in Dardic languages.

One of the major Dardic innovations is the (partial) loss of aspiration, mainly in voiced stops/obstruents (e.g., most Dardic languages, except Torwali, which has both voiced and voiceless aspirated stops), but sometimes also in voiceless obstruents (e.g., Pashai and Grangali). Loss of aspiration is a recent development in Dardic and could be a result of contact with Iranian languages where aspiration was lost at a very early stage. Traces of aspiration in Dardic are sometimes observed in the development of tonal contrasts (e.g., Khowar *buúm* ‘earth’ vs. IA *b^huumi*; Pashai *duū^m* ‘smoke’ vs. IA *d^hūvāā* and OIA *d^huumra*).

Another innovation of Dardic languages is the development of retroflex affricates ɟ, ɟ^h, ʝ, and ʒ from various OIA clusters. This change could also possibly be attributed to areal influence. Retroflex affricates are also found in Burushaski spoken in the northwest frontier province in Pakistan and in Dravidian languages (It is a well-established theory that Dravidians were the original inhabitants of the region before the advent of Aryans who pushed Dravidians down south. The assumption is further strengthened by the presence of Brahui, a Dravidian language, in Afghanistan). Dardic languages have also developed a contrast between voiceless and voiced fricatives (e.g., *s/z* and sometimes *x/ɣ*), a distinction absent in most NIA as well as OIA languages but present in the Iranian languages. The vowel systems of many Dardic languages have undergone several changes.

Vowel inventories as large as the 16-vowel system of Kashmiri or the 20-vowel system of Kalasha are an example. Some of the phonological changes with respect to the vowels are vowel epenthesis, consonantal palatalization, and vowel harmony.

Morphosyntax

Like most areal languages, Dardic languages are typically postpositional with S(ubject)-O(bject)-V(erb) word order. The only exception, however, is Kashmiri, which is a V2 language (i.e., the inflected verb occurs at clause-second position). Most languages exhibit Split-Ergative case marking (e.g., Dameli, Gawarbuti, Grangali, Pashai, Kalam Kohistani, Kashmiri), except a few that are Nominative-Accusative (e.g., Khowar and Kalasha) or fully Ergative (e.g., Shina). Complementizers in most Dardic languages are derived from the verb 'say' (e.g., Kalasha, Khowar, Palula, and Shina), but in many others *ki/ke* (*ki/zi* in Kashmiri), also used in most contact languages, is employed as a complementizer. Relative clauses are mostly prenominal with a fully finite verb, sometimes without a relative pronoun, and a relative-correlative construction – a typical IA and areal syntactic feature. Overtly marked case-endings behave like postpositions. Nominals preceding postpositions appear in oblique case (another typical areal feature).

Agreement patterns vary across languages. Both subject and object pronominal clitics may appear on the inflected verb (e.g., Kashmiri). In many Dardic languages animacy has become grammaticized (e.g., Khowar, Kalasha, and Torwali). Feminine gender is often marked by consonantal palatalization (e.g., Pashai, Shumashti, and Kashmiri). Most Dardic languages have developed a vigesimal counting system with (10 + n) numeral structure (sometimes with modifications) as compared to the typical IA (n + 10) system. Kashmiri is an exception, with the IA (n + 10) numeral system. A significant morphological feature of Dardic languages is a three-term (or larger), instead of the typical two-term, deictic system. For instance, the three-fold demonstrative systems of Pashai (proximate *yo* 'this', distal *e-lo* 'this', remote (*e*)-*se* 'that'; Bashir, 2003: 828) and Kashmiri (proximate *yi* 'this',

visible *hu/ho* 'that; masculine/feminine', invisible/remote *su/so* 'that; masculine/feminine').

Bibliography

- Baart J L G (1997). *The sounds and tones of Kalam Kohistani*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.
- Bashir E (2003). 'Dardic.' In Cardona G & Jain D (eds.) *The Indo-Aryan languages*. London/New York: Routledge, Taylor & Francis Group. 818–894.
- Grierson G A (1906). *The Pisaca Languages of north-western India*. London: The Royal Asiatic Society.
- Grierson G A (1919). *The Linguistic Survey of India*, 8, II. Calcutta: Superintendent Government Printing, India.
- Hallberg D G (1992). 'The Languages of Indus Kohistan.' In Rensch *et al.* (eds.). 83–141.
- Kachru B B (1969). 'Kashmiri and other Dardic languages.' In Sebeok T A (ed.) *Current trends in linguistics 5, Linguistics in South Asia*. Paris: The Hague. 284–306.
- Koul O N (2003). 'Kashmiri.' In Cardona G & Jain D (eds.) *The Indo-Aryan languages*. London/New York: Routledge, Taylor & Francis Group. 895–952.
- Masica C P (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Mohiuddin A (1998). *A fresh approach to the history of Kashmir*. Srinagar: Book Bank.
- Morgenstierne G (1965). 'Dardic and Kafir languages.' In Lewis B, Pellat C & Schacht J (eds.) *Encyclopedia of Islam 2* (new edition). Leiden: E. J. Brill. 138–139.
- Radloff C F (1992). 'Dialects of Shina.' In Backstrom P C & Radloff C C (eds.) *Languages of northern areas, sociolinguistic survey of northern Pakistan 2*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics. 89–203.
- Rensch C R (1992). 'Patterns of language use among Kohistanis of the Swat valley.' In Rensch *et al.* (eds.). 3–62.
- Rensch C R, Decker S J & Hallberg D G (1992). *Languages of Kohistan, sociolinguistic survey of northern Pakistan 1*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.
- Schmidt R L (1988). 'Paálus/kostyó/ 'Shina revisited.' *Acta Orientalia* 59, 106–149.
- Sharma D D (1998). *Studies in Tibeto-Himalayan languages 6: Tribal Languages of Ladakh, I*. New Delhi: Mittal Publications.

Dhivehi

J W Gair, Cornell University, New York, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

General

Dhivehi (Dhivehi Bas, Divehi, Maldivian) is the language of the Maldives, where it is the official language, with approximately 3.2 million speakers (U.N., 2003). It is also spoken by about 3000 inhabitants on the island of Minicoy (Maliku), a territory of India, where it is known as Mahl. It is an Indo-European language of the Indo-Aryan family, and its closest relative is Sinhala of Sri Lanka, with which it forms a separate southern (island) subbranch. Though the two languages are clearly related and share many structural features, they are mutually unintelligible. The manner and date of their separation is uncertain, and serious scholars have proposed widely varying times. It has been suggested, on the one hand, that they represent a common source but separate settlements in the mid-first millennium B.C.E., the generally recognized date for the arrival of Sinhala in Sri Lanka. On the other hand, a date as late as the 10th century through the importation of Sinhala into the Maldives has been proposed. One problem is that some important sound changes that would appear to be common to the two, when examined closely, turn out to have slightly different conditions. Thus, there are signs of divergence as early as the 1st century B.C.E., but these are followed at several subsequent points by changes shared by the two languages that are of a kind that are uncommon elsewhere. Certainly the earliest and latest dates proposed seem extreme on the basis of more recent research, and one scenario might be that divergence began around the first century B.C.E. but was followed by Sinhala influence over time, together with some dialect admixture within Dhivehi and contact of both languages with South Indian Dravidian (for a detailed account see Cain, 2000).

The base vocabulary of Dhivehi is Indo-Aryan, but it has incorporated many words from other languages, including Arabic, English, and Dravidian as well as Sinhala.

The Maldives are a chain of over 1000 islands in atolls (a word borrowed from Dhivehi) ranging 450 miles north and south, and there are significant dialect differences within it. The standard language is based on the language of Malé, the capital, in the North. The speech of the southernmost atolls differs from the standard in important respects. There is also differentiation within the southernmost atolls (see Fritz, 2002). The Mahl of Minicoy is mutually intelligible with the Malé variety, and there

is significant cultural interaction between Minicoy and the Maldives. Maldivian literacy is high: almost 99% in 2001.

Although Dhivehi is the official language of the Maldives, and the first language of the regular inhabitants of the islands, there is widespread knowledge of English, and English is the medium of instruction in government schools.

Phonology

Like other Indo-Aryan languages, Dhivehi has voiced and voiceless consonants and a contrast between dental and retroflex stops. There are five vowels: *i*, *e*, *a*, *o*, *u*, that occur long and short. A retroflex grooved spirant /ʃ/ is unique to Dhivehi and derives from intervocalic retroflex /ɖ/, with the latter reintroduced through loanwords. Two notable features, shared with Sinhala, are the lack of any aspirated consonant series and a set of prenasalized stops /^mb, ⁿd, ^ŋg/ that contrast with nasal-stop clusters. Unlike in Sinhala, the consonant /f/ is of quite frequent occurrence, having arisen from a sound change /p/ > /f/, as well as from loanwords.

Orthography

The current Dhivehi script, known as ‘Thaana,’ is unique to that language. It is written left to right, and the characters are made to resemble Arabic, reflecting the influence of Islam. They are not Arabic, however, although the first nine letters were fashioned after Arabic numerals. The system is of the alphasyllabic type, with all consonants and vowels being written, but grouped in syllabic clusters. Vowels following consonants are written above or below them, as in many South Asian scripts, but unlike most Indic scripts, consonants do not imply an unmarked inherent vowel. Also, initial or independent vowels do not have their own signs but are written using the character *alifu* (ا) which has no phonetic value by itself, but serves as a vehicle for the same vowel diacritics that are used with consonants. Consonants not followed by a vowel are marked with *sukun* (◌◌). Thus, the name of the language in Thaana, with a transliteration (read right to left) and phonological representation (left to right) is:

= <s - sukun> + <ba> <hi> + <ve>

 <dhí> = /divehi bas/.

(For a fuller account, see Gair and Cain, 2000). The basic Thaana alphabet is supplemented by a set of characters for the numerous Arabic borrowings.

There is also an official romanization, designed to avoid diacritics, that reflects some English influence. Thus, long *a* is written <aa>, long *e* is <ey>, long *o* is <oa>, and retroflex *ṣ* is <sh>. The combinations <th> and <dh> represent dental stops, as in *Dhivehi* and *Thaana*. Thus the <h> represents dental articulation (vs. retroflex), not aspiration, which is lacking in Dhivehi. (For a description, see Maniku and Disanayake, 1990 or the Thaana equivalents in Cain and Gair, 2000.)

Morphology

Although there are differences in inflectional forms across dialects, the major categories are shared (see Fritz, 2002; Cain and Gair, 2000).

Nouns are human or nonhuman in gender. They inflect for case (direct, dative genitive, instrumental, and locative), definiteness (definite, indefinite, and unspecified), and number (singular and plural, though the singular is generally unmarked in inanimate nouns).

The verbal system includes derivational relationships in stems between active, involitive/intransitive, and causative forms. There are finite and nonfinite forms based on three stems: present, past, and participial. Finite verbs inflect for three tenses (past, present, and future) and three aspects (habitual, present, and perfect). Person and number categories vary with the dialect.

Syntax

The basic constituent order in Dhivehi is subject–object verb, as illustrated in (1), though other orders are possible for pragmatic effect. It is a thoroughgoing ‘right headed’ language, with complements and modifiers preceding heads, as illustrated in (2). This includes relative clauses, which use a tensed relativizing (adjectival) form of the verb, as in (3). Note that, the singular of inanimate nouns is commonly used in the plural as well. An important feature of Dhivehi, shared with Sinhala, is the presence of a focus cleft construction of frequent occurrence in discourse. In Dhivehi, this is formed with a ‘prefocus’ form of the verb, as in (4).

- (1) alī e mīhā duṣ.
Ali that person see.PAST
‘Ali saw that person.’

- (2) mi raⁿgaḷu tin fot
this good three book-Sg/Pl
‘these three good books’

- (3) [hassan alīy-aṣ din] fot
Hassan Ali-DAT give.PAST.REL book-Sg/Pl
‘the book that Hassan gave Ali’

- (4) aharen danī e avaṣaṣ.
I go-PRE.FOC that neighborhood-DAT
‘It is to that neighborhood that I am going.’

Bibliography

- Bell H C P (1922–35/1998). *Excerpta maldiviana*. New Delhi: Asian Educational Services.
- Bell H C P (1940/1985). *The Maldive Islands: monograph on the history, archaeology, and epigraphy*. Colombo: Ceylon Government Press. Malé, Maldives: National Centre for Linguistic and Historical Research.
- Cain B D (2000). *Dhivehi (Maldivian): a synchronic and diachronic study*. Ph.D. diss., Cornell University.
- Cain B D & Gair J W (2000). *Dhivehi (Maldivian). Languages of the world/materials* 63. Munich: Lincom Europa.
- De Silva M W S (1969). ‘The phonological efficiency of the Maldivian writing system.’ *Anthropological Linguistics* 11.7, 199–209.
- De Silva M W S (1970a). ‘Some affinities between Sinhalese and Maldivian.’ *Journal of the Royal Asiatic Society (Ceylon Branch)*, n. s. 14, 202–207.
- De Silva M W S (1970b). ‘Some observations on the history of Maldivian.’ *Transactions of the Philological Society*, 1970. 137–162.
- De Silva M W S (1979). *Sinhalese and other island languages in South Asia. Ars Linguistica* 3. Tübingen: Gunter Narr Verlag.
- Disanayake J B (1986). ‘Maldivian and Sinhala: some phonological observations.’ *Journal of the Royal Asiatic Society (Sri Lanka)* 30, 81–100.
- Fritz S (2002). *The Dhivehi language: a descriptive and historical grammar of Maldivian and its dialects*. (2 vols). *Beiträge zur Südasiensforschung; Südasiens-Institut Heidelberg*. Heidelberg: Ergon Press.
- Gair J W & Cain B D (1996). ‘Dhivehi writing.’ In Daniels P T & Bright W (eds.) *The World’s writing systems*. New York and Oxford: Oxford University Press. 564–568.
- Geiger W (1902). ‘Etymological vocabulary of the Maldivian language.’ *Journal of the Royal Asiatic Society, Great Britain and Ireland (3rd series)* 909–938.
- Geiger W (1919/1996). *Maldivian linguistic studies*. Willis J C (trans.) & Bell H C P (ed.) *Journal of the Ceylon Branch of the Royal Asiatic Society* 27 (extra number). New Delhi: Asian Educational Services).
- Karunatilake W S (2000). *Historical phonology of Sinhalese: from Old Indo-Aryan to the 14th century A.D.* Colombo: S. Godage and Brothers.
- Maloney C (1980). *People of the Maldive Islands*. Madras: Orient Longman.
- Maniku H A & Disanayake J B (1990). *Say it in Maldivian (Dhivehi)*. Colombo: Lake House Publishers.
- Reynolds C H B (1978). ‘Linguistic strands in the Maldives.’ In Maloney C (ed.) *Contributions to Asian Studies, vol. II: language and civilization change in South Asia*. Leiden: E. J. Brill. 156–166.
- Reynolds C H B (1993). *Maldives. World bibliographical series volume* 158. Oxford: Clio Press.
- Reynolds C H B (2003). *A Maldivian dictionary*. London: Routledge/Curzon.

Diachronic Morphological Typology

A Harris and Z Xu, State University of New York, Stony Brook, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

The morphology of various languages has been typed in terms of prefix, suffix, infix, and circumfix, although this omits some types, such as ablaut and subtractive morphology. This article considers the diachronic origins of these affix types and the diachrony of the so-called suffixing preference. Languages have also been typed as fusional, agglutinating, or isolating; the diachrony of these and related concepts is explored in the last section.

Diachronic Origins of Affixes

Prefixes and Suffixes

Most affixes are formed either from old affixes or from grammatical words through the process of grammaticalization. For example, the Old Georgian ergative case suffix, *-man*, is derived from the ergative singular case form of the definite article, *man* ‘the,’ which followed the noun and was unstressed. In Modern Georgian the ergative suffix is *-ma* or, after a vowel, *-m*. Prefixes are formed in a similar way, usually from grammatical material proclitic to a lexical item.

Infixes

An infix is a morpheme appearing within another morpheme. Alan C. L. Yu showed that typologically there are several phonological units to which infixes can be adjacent: first consonant, first vowel, final syllable, final vowel, stressed syllable, stressed vowel, stressed foot (Yu, 2003). Yu argued that the reason why an infix tends to be close to the boundary of the base to which it attaches is because it has been either a prefix or suffix historically. For example, the Proto-Muskogean plural **oho-* prefix developed into a prefinal syllable infix, *-ho-*, in Creek-Seminole (Muskogee) (e.g., *likw-i*: ‘rotten’ → *likhow-i* ‘rotten.PL’) (Martin, 1994, cited in Yu, 2003).

Yu discussed four processes that give rise to infixes: entrapment, metathesis, reduplication mutation, and prosodic stem association. For example, in Proto-Muskogean the mediopassive proclitic **il-* appeared after the applicative **a-* and the plural **oho-*, which were later reanalyzed as part of the verb stem, entrapping the intervening affix **il-* (e.g., **oho-il-icca* ‘be shot’ → Alabama *holicca* ‘be shot’) (Martin and Munro, 2005, cited in Yu, 2003). A case of metathesis causing infixation comes from Copainalá Zoque (CZ). The third-person marker was historically a

prefix, **i-*, which later became a glide. Metathesis turned all **j + C(onsonant)* sequences into *Cj* in CZ (e.g., *mula* ‘mule’ → *mjula* ‘his mule’) (Yu, 2003: 221). Reduplication mutation can also trigger infixation. In Pre-Trukic the durative form of **kasam^wó:nu* ‘pay chiefly respect to’ was **kak-kasam^wó:nu* ‘be in the habit of paying chiefly respects to.’ After the dropping of initial **k*, the reduplicated form became **ak-kasam^wó:nu*, which was later reanalyzed as **akk-asam^wó:nu*. The V(owel)kk-affix was generalized in other vowel-initial verbs. After a process took place in which a glide *w* was inserted verb-initially and became part of the base, *Vkk*-turned into an infix. For example, in Trukese the durative form of *win* ‘drink’ is *w-ikk-in* ‘be in the habit of drinking.’ A case of prosodic stem association causing infixation is what Yu called “Homeric infixation” or “*ma*-infixation,” a colloquial expression indicating “roughly attitudes of sarcasm and distastefulness” (Yu, 2003: 249). For example, *-ma-in whatchimacallum* ‘what you may call him’ comes from the word ‘may.’ Because the construction with *-ma-* indicates colloquialism and *-ma-* usually appears between two metrical feet (e.g., [*whatcha*]-*ma*[*callit*], [*thingu*]*ma*[*bob*]), speakers analyze it as an infix.

Circumfixes

A circumfix is a single complex affix composed of a prefixal and a suffixal part functioning together. The circumfix as a whole expresses a single meaning or category, and the parts of a circumfix are not affixed separately to the same base to which they are affixed together.

The Proto-Austronesian circumfix **ka-an* has two meanings, only one of which is described here. In some of the Formosan languages, and in languages of the Philippines and western Indonesia, *ka-an* forms nouns, often ones meaning a location or having some other concrete meaning. Proto-Austronesian also had a prefix **ka-*, which formed nouns, and a suffix **-an*, marking ‘locative focus.’ Thus, two Proto-Austronesian affixes may have come together, combining their meanings; these became a circumfix and no longer function as a distinct prefix and suffix, even though the prefix or the suffix is also continued in some languages (Blust, 2003).

We learn more by looking at the facts of Algic languages. In Algonquian, circumfixal pronominals are used in noun possession and the independent order, the set of verbal paradigms used in most main clause statements. These are reconstructed as follows:

- | | |
|--------------------|------------|
| (1) 1 PL exclusive | *ne-enaan- |
| 1 PL inclusive | *ke-enaw- |
| 2 PL | *ke-waaw- |
| 3 PL | *we-waaw- |

Proto-Algonquian *-enaan- (first person exclusive) is cognate to Wiyot *binód* ‘we, us,’ a separate word; Proto-Algonquian *-waaw- is cognate to Wiyot *wow*, a particle or postposition used only to pluralize the second person pronoun, *khil*. Wiyot is a language of the Ritwan group, distantly related to Algonquian, the two groups together forming the Algic family. Thus, it appears that in Proto-Algonquian the independent words *-enaan and *-waaw came to be suffixed to forms already prefixed with *ne- and *ke- and that at some point these prefixal and suffixal pairs were reanalyzed as circumfixes (Ives Goddard, personal correspondence).

Summary

Prefixes and suffixes are most often derived either from old affixes, not illustrated here, or from grammatical words, usually with an intermediate stage as a clitic. Infixes are usually derived from prefixes or suffixes, and circumfixes from a combination of prefixes and suffixes. Usually affixes occur in the same position, relative to a base, as the clitic from which they derive.

The Suffixing Preference

The suffixing preference, first noted by Edward Sapir (Sapir, 1921), is the generalization that suffixing is far more frequent than prefixing crosslinguistically.

John A. Hawkins and Gary Gilligan set out to quantify the suffixing preference (Hawkins and Gilligan, 1988). Using a sample of about 200 languages, they studied the distributions of specific affixes of both nouns and verbs (e.g., case markers, tense markers) for distribution as prefixes or suffixes in relation to word order type (head final vs. head initial). Their findings show a strong suffixing preference for affixes of most types tested (but a slight prefixing preference for object markers and only a weak suffixing preference for markers of subject, negation, and voice). They state 18 implicational universals, several of them exceptionless, such as the first one below:

- If a language has case affixes on n[oun], they are always suffixed.
- If a language has SOV, causative affixes on v[erb] (if any) are suffixed with more than chance frequency.

Hawkins and Gilligan proposed an explanation of the suffixing preference in terms of competition between two forces. One is an independent

head-ordering principle (HOP), which states that “heads are identically ordered relative to their modifiers at [morphological and syntactic] levels” (Hawkins and Gilligan, 1988: 220). The second component of the competition is a processing preference for ordering the important part of a word before the less important, and hence for stems before affixes; this is elaborated in Hawkins and Cutler (1988).

Christopher Hall observed a problem with the proposed explanation in terms of the HOP and processing preferences: it is incomplete in that it does not explain how processing preferences are linked to language types (Hall, 1988). That is, the notion of competition between the HOP and processing preferences provides no means of implementing these preferences. Hall proposes instead that the processing mechanism influences the diachronic changes that lead to the formation of prefixes or suffixes. Thus, suffixing is more common because the language processor influences diachronic processes that determine whether an affix will develop as a prefix or suffix.

Affixes in Verb Forms

Most grammatical affixes in verbs are derived from old verbal morphology, from auxiliaries, or from adverbs; new subject and object agreement affixes may originate as pronouns or as auxiliaries. In all instances, independent words usually become clitics before they go on to become affixes. It is widely assumed that the position of an affix is usually due to the position of the clitic from which it derives. This assumption is often attributed to Givón, 1971, but in fact it was made traditionally in historical grammar. Many linguists believe that the historical origin of an affix is one part of the explanation of its position as a prefix or suffix and thus that history plays an important part in explaining the suffixing preference.

Joan M. Bybee, William Pagliuca, and Revere Perkins studied the suffixing preference by surveying 71 languages (Bybee *et al.*, 1990). They observed that grammatical morphemes (e.g., adpositions, clitics, particles, and auxiliaries) following the stem are more likely to become affixes than are grammatical morphemes preceding the stem. Categorizing the languages in their sample into three groups based on word order type, i.e., V(erb)-initial languages, V-medial languages, and V-final languages, they distinguished bound and unbound grammatical morphemes in both preposed and postposed positions. They focused on the relation between verb stems and grammatical morphemes, specifically particles and auxiliaries.

Their data showed that postposed grammatical morphemes far outnumber preposed ones and have

a strong tendency to be bound. In the 32 V-final languages they studied, Bybee, Pagliuca, and Perkins found 1018 postposed grammatical morphemes, 80% of which are bound (i.e., are suffixes), and only 233 preposed grammatical morphemes. Based on the hypothesis that affixes that develop from lexical or grammatical forms stay in their original positions, they explain that in V-final languages auxiliaries, which typically follow main verbs and are a primary source of grammatical morphemes, become suffixes. Additionally, their data suggest that it is more usual for person/number markers to be suffixes in V-final languages (see Table 1). Table 1 shows that the number of postposed person/number markers (171) is almost twice that of preposed person/number markers (90), and all postposed person/number markers are bound (are suffixes) in languages of this type.

Putting aside the cases in which person/number markers postpose in V-initial languages, Bybee, Pagliuca, and Perkins observed no suffixing preference in V-initial languages. “The tendency for preposed grams [grammatical morphemes] to be bound [is] slightly stronger than the tendency for postposed grams to be bound, but ... it is statistically not significant” (Bybee *et al.*, 1990: 13) (see Table 2).

Turning now to V-medial languages, Bybee, Pagliuca, and Perkins found that they show a slight

preference for preposing grammatical material, together with a strong preference for these morphemes not to be bound (see Table 3). Their account of this dispreference for prefixing in V-medial languages is that clause-internal auxiliaries do not necessarily attach to verb stems; for example, they may “fuse with pronouns and with one another to form an auxiliary complex that occurs between the subject and the verb, or in second position in the clause, without ever fusing with the verb” (Bybee *et al.*, 1990: 31); prefixing is conditioned by the semantic relevance of clause-internal grammatical morphemes to verb stems.

On the basis of data from 237 languages, Anna Siewierska and Dik Bakker found that there is no suffixing preference for agreement markers (Siewierska and Bakker, 1996). They attributed this in large part to the inclusion in their sample of many languages of North America, a large number of which have subject or object agreement prefixes. However, even without the languages of North America, their data showed a slight preference for subject agreement prefixes, and only a slight preference for object agreement suffixes. Thus, they found no suffixing preference for agreement.

Studies that show us what happens to particular grammatical morphemes over time are especially valuable in understanding these issues. Comrie (1980) showed that Classical Mongolian (attested from the 13th century) had the word orders SOV and adjective-noun and lacked verb agreement. When the subject was a pronoun, the language permitted a variant of SOV, with an unstressed pronoun subject following the verb. The permitted orders in Classical Halh Mongolian.

- (2a) bi med-ne
I know-PRES
‘I know’
- (2b) med-ne bi
know-PRES I
‘I know’ (Comrie, 1980: 90)

Table 1 Position and boundedness of person/number markers in V-final languages

| | Nonbound | Bound | All |
|-----------|----------|------------|-----------|
| Preposed | 13% (10) | 87% (80) | 35% (90) |
| Postposed | 0 | 100% (171) | 65% (171) |

Reproduced from Bybee, Pagliuca, and Perkins, 1990. ‘On the asymmetries in the affixation of grammatical material’, in *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on his 75th Birthday*, ed. by Croft, Denning, and Kemmer, 1–42. Amsterdam: John Benjamins 9. With kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia www.benjamins.com.

Table 2 Position by boundedness for nonperson/number grammatical morphemes in V-initial languages

| | Nonbound | Bound | All |
|-----------|----------|----------|----------|
| Preposed | 19% (13) | 81% (57) | 53% (70) |
| Postposed | 27% (17) | 73% (46) | 47% (63) |

Reproduced from Bybee, Pagliuca, and Perkins, 1990. ‘On the asymmetries in the affixation of grammatical material’, in *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on his 75th Birthday*, ed. by Croft, Denning, and Kemmer, 1–42. Amsterdam: John Benjamins 13. With kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia www.benjamins.com.

Table 3 Position with respect to verb by boundedness for the 31 V-medial languages

| | Nonbound | Bound | All |
|-----------|-----------|-----------|-----------|
| Preposed | 60% (298) | 40% (200) | 54% (498) |
| Postposed | 19% (82) | 81% (341) | 46% (423) |

Reproduced from Bybee, Pagliuca, and Perkins, 1990. ‘On the asymmetries in the affixation of grammatical material’, in *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on his 75th Birthday*, ed. by Croft, Denning, and Kemmer, 1–42. Amsterdam: John Benjamins 6. With kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia www.benjamins.com.

Buriat and other daughters have subject agreement suffixes derived during the historical period from the variant in (2b). It is usually assumed that independent pronouns occur in the same basic position as full nouns, unless the pronouns cliticize, and this is recognized as the first stage on the way to grammaticalization as an agreement marker. What we do not know in this case is whether in a stage of Mongolian before the first attestation, free pronouns occurred only where full noun subjects occur, as usually assumed crosslinguistically, and then moved into enclitic position. If so, this trivializes the idea that we can explain the order in which affixes occur by the order of the words from which they derive.

We do know that in the history of French, object pronouns occurred after the verb in independent forms and later, when cliticized, moved into immediately preverbal position. SVO order, with object pronouns following the verb, was established in Latin as early as the 5th century. In Old French (SVO), stressed pronouns remained in the position used for full NPs, while unstressed pronouns were attracted to the verb, usually occurring immediately before it. According to many, these Old French clitics have developed into agreement prefixes in the modern language, though others consider them still to be proclitics (and they are still written as separate words). In either case, this is another example of a clitic developing in a position different from that of the stressed pronoun.

It is widely agreed that the position of an affix is very often determined by the position of its etymon before grammaticalization. New affixes most often develop from clitics, but the position of a clitic is largely determined by prosodic conditions. As discussed above, a pronoun may be cliticized in a position in which it did not occur before becoming a clitic, and a definite article may also cliticize in a new position. For prosodic reasons, clitics are attracted to certain positions relative to other words or relative to the clause as a whole. It is the position of a clitic, not that of the lexical item to which it corresponds, that largely accounts for the position in which it grammaticalizes as an affix.

Bybee (1988: 358) and Bybee *et al.* (1990) suggested that the apparently lower tendency of preposed grammatical material (in comparison with postposed) to grammaticalize as an affix may be due to the fact that other words intervene. Recent work on grammaticalization in progress seems to support this. Kraehenmann and Plank (forthcoming) shows that proclitic articles on their way to becoming prefixes in two German dialects seem to be prevented from doing so by the intervention of other lexical items, such as adjectives, between the article and

noun. For additional discussion of the relation between order of words and order of morphemes from a historical point of view, see Harris and Campbell (1995: 199 ff.).

Affixes in Nominal Forms

When we turn to the issue of the suffixing preference in nouns, the facts are a little different. There are very few case prefixes in the languages of the world, and even where they do occur, it appears that there are no whole declensions consisting just of case prefixes.

There are several known sources for affixal case markers, including old case markers, definite articles, and adpositions. Greenberg (1978) showed that definite articles, themselves derived from demonstrative pronouns, often become gender class markers or case markers. For example, certain case suffixes can be shown to be derived from postposed definite articles in Georgian (see 'Prefixes and Suffixes' above), and whole new declensions may also be constructed in this way.

Postpositions often become case suffixes: for example, Votic (Vod), a Balto-Finnic language, still retains the comitative postposition *kāsa*, *kaaza*, and from it has developed a case suffix shown in *mineka* 'with what,' *jummalaga* 'with God,' *lehmika* 'with cows' (Oinas, 1961: 36–40).

In Hungarian, inflected forms of nouns have been reanalyzed as complex cases. It appears, however, that these inflected nouns went through an intermediate stage as complex postpositions before becoming suffixes. Similarly, serial verbs may become case prefixes, but it is most likely that they are prepositions at an intermediate stage. Thus, we cannot be sure that nouns and serial verbs constitute distinct sources of cases; both are known to be sources of adpositions, and it is most likely that it is these that give rise to cases.

Greenberg (1978) documents in detail that in various languages, mostly African, the definite article can be preposed or postposed and may become a gender prefix or suffix respectively. As we have seen, definite articles may instead be grammaticalized as case markers, but virtually all of the latter are suffixes.

There is a great deal here we do not yet understand. In languages that have definite articles that precede nouns, do the articles not grammaticalize as case markers, or do they become enclitics and then grammaticalize as suffixal case markers? Whatever historical behavior results in only case suffixes, why does it apply to case marking and not to gender marking (or why does it apply to gender marking to a lesser extent)?

Summary

Research has confirmed the existence of a suffixing preference in both nouns and verbs. For V-final languages this preference is very strong, but for both V-initial and V-medial languages suffixes are preferred in the verb less strongly than in V-final languages. There are also significant differences among different types of verbal affixes, with agreement markers more likely than most other types to be prefixes, looking at languages of all types together. There is also evidence that in at least some language types postposed grammatical material is more likely to be bound than is preposed grammatical material. In the noun we find an overwhelming preference for case suffixes; other noun affixes, such as markers of gender and number, exhibit a somewhat weaker preference for suffixal position.

The position of affixes is best explained historically and with reference to language processing. Historical explanation of affix position refers to the position of the clitics from which prefixes and suffixes derive, and to the additional events that result in the creation of infixes and circumfixes. There is still much to be learned about the positions in which clitics form diachronically.

Fusional, Agglutinating, Isolating

A morphological typology to which linguists have returned repeatedly designates languages as fusional, agglutinating, or isolating. A language is said to be fusional (or flectional or inflecting) if the separation between morphemes is not readily apparent. Characteristically, in such languages inflectional morphemes each express two or more categories (for example, number + case in the noun, or tense + person + number in the verb). Many Indo-European languages, such as Latin, Greek, and Sanskrit, are examples of this type. In Russian declensions, as shown in (3), each suffix indicates number and case, and some also indicate declension class.

| | | | | |
|---------------|----------|----------|-----------|------------|
| (3) | I | | II | |
| | Singular | Plural | Singular | Plural |
| Nominative | stol | stol-y | komnat-a | komnat-y |
| Accusative | stol | stol-y | komnat-u | komnat-y |
| Genitive | stol-a | stol-ov | komnat-y | komnat |
| Dative | stol-u | stol-am | komnat-e | komnat-am |
| Instrumental | stol-om | stol-ami | komnat-oi | komnat-ami |
| Prepositional | stol-e | stol-akh | komnat-e | komnat-akh |
| | 'table' | | 'room' | |

If we consider here, for example, the genitive plural of *stol*, *stol-ov*, we cannot say that one part of *-ov* indicates the genitive, another the plural, and a third the first declension. Rather, *-ov* as a whole indicates all three of these values.

In an agglutinating language, words are made up of a sequence of morphs, each expressing a separate category, and the boundaries between morphemes are unambiguous. Turkish is often cited as an example of an agglutinating language, for here number and case are expressed by different morphemes in the noun, for example. Old Georgian (4) illustrates agglutination in the verb.

(4) Old Georgian 'write' in the optative

| | Singular | Plural |
|---|----------|------------|
| 1 | v-c'er-o | v-c'er-o-t |
| 2 | s-c'er-o | s-c'er-o-t |
| 3 | c'er-o-s | c'er-o-n |

In the Old Georgian verb paradigm in (4), the morpheme *v-* marks first-person subjects without respect to number, while *s-* marks second-person subjects. The optative is indicated by the suffix *-o*, and plurality of the subject by *-t*. These aspects of this paradigm are agglutinative, but some fusion is found in the third person, where one suffix, *-s*, combines the marking of person and singular number, while *-n* combines the marking of person and plural number.

In an isolating language the forms of words are invariable, and such languages are sometimes said to have no morphology. An isolating language depends more upon syntax to express grammatical categories and relationships among words. Vietnamese is often cited as an example; see (5).

- (5) Khi tôi đến nhà bạn tôi,
 when I come house friend I
 chúng tôi bắt đầu làm bài.
 PL I begin do lesson
 'when I came to my friend's house, we began to
 do lessons'
 (Comrie, 1988: 40)

Each morpheme in (5) is a word, with the possible exception of *bắt đầu* 'begin.' There is no morphological variation for either tense or case. Plurality is indicated by the addition of a separate word *chúng*.

This typology can usefully be broken down into two scales, each with a single criterion – the index of synthesis (based on the number of morphemes per word) and the index of fusion (based on the number of categories expressed per morpheme). These indices represent an advance, but to the extent that they require generalizing over an entire language, they are still abstract and subject to differing interpretations. That is, many languages mix types, even within a single paradigm (as illustrated in [4]), and, for this reason and others, two specialists could still reach different conclusions about the type to which a given language belongs.

When we look at historical aspects of this typology, we find that many languages or language families have changed type historically. An isolating language can become agglutinating by accumulating new affixes through one of the means described in 'Diachronic Origins of Affixes.' In an agglutinating language, affixes may merge, and the language may then become fusional. A fusional language may lose its affixes, becoming an isolating language. For example, Old English had fusional markers in both the noun and the verb and has lost most of these, becoming more isolating (for additional examples, see Crowley, 1992: 134–136). Although it is sometimes implied that change cycles only in this direction – isolating → agglutinating → fusional → isolating, etc., – it is possible for a language to change in the opposite direction as well. An example of change from agglutinating in the direction of isolating is the Artašen dialect of Laz, which has lost all of its case markers, though it has retained most of its other morphology (Harris and Campbell, 1995: 216–217). Agglutinative affixes can be lost as well as fusional ones. Notice, too, that if a language goes in the direction specified but skips a step, that is equivalent to going in the opposite direction.

Closely related to this typology are the general labels 'synthetic' and 'analytic.' A construction, is said to be synthetic if several categories are expressed within a single word, while in an analytic construction, categories are expressed through periphrasis, that is, by combining words in a phrase. English *jumps* is synthetic, while *is jumping* is analytic because *jumping* is combined with *is*. In the history of French we can see a cycle from synthetic constructions in Classical Latin to analytic constructions and back to synthetic, for at least some forms. For example, the third person singular form of the imperfective future of the verb 'do, make' in Classical Latin was *faciet* 'he will do,' which in Vulgar Latin was replaced with the analytic (or periphrastic) expression *facere habet*, where the infinitive of 'do,' *facere*, is combined with an inflected form of the auxiliary 'have,' *habet*. In Old French this emerges as the single word *fera*, the third person singular of the future (Harris, 1978). In many ways this is a classic example of the synthetic-analytic cycle, but note that although the language moves from fusional to somewhat isolating and back to fusional, no agglutinative stage is involved. In a truly agglutinative verb form, person and number would be expressed by distinct morphemes, as in the Georgian example above, whereas at all stages of this change person and number are expressed together. When an auxiliary expresses person and number in a single morpheme, a natural outcome is for these categories to

continue to be expressed in a single morpheme in the synthetic form derived from the auxiliary. A similar complete cycle can be seen in Egyptian, where Old Egyptian was synthetic, Middle and Late Egyptian analytic, and Coptic, its descendant, again synthetic (Hintze, 1947; Hodge, 1970).

The classifications fusional, agglutinating, and isolating provide very general ways of typing languages; not all languages fit well into any one of these types, since a language may combine types in various ways. Languages may change in the direction agglutinating → fusional → isolating → agglutinative etc., but changes in the opposite direction are also known. Cyclic changes are attested, but it is not clear that these necessarily include all types.

Summary

A historical approach, sometimes paired with insights from language processing, provides an explanation of typological phenomena, including the occurrence of morphemes of different types (prefixes, suffixes, infixes, and circumfixes) and the suffixing preference.

Bibliography

- Blust R (2003). 'Three notes on early Austronesian morphology.' *Oceanic Linguistics* 42, 438–478.
- Bybee J L (1988). 'The diachronic dimension in explanation.' In Hawkins (ed.). 350–379.
- Bybee J L, Pagliuca W & Perkins R D (1990). 'On the asymmetries in the affixation of grammatical material.' In Croft W, Denning K & Kemmer S (eds.) *Studies in typology and diachrony: papers presented to Joseph H Greenberg on his 75th birthday*. Amsterdam: John Benjamins. 1–42.
- Comrie B (1980). 'Morphology and word order reconstruction: problems and prospects.' In Fisiak J (ed.) *Historical morphology*. The Hague: Mouton. 83–96.
- Comrie B (1988). *Language universals and linguistic typology*. Chicago: University of Chicago Press.
- Crowley T (1992). *An introduction to historical linguistics* (2nd edn.). Oxford: Oxford University Press.
- Givón T (1971). 'Historical syntax and synchronic morphology: an archaeologist's field trip.' *Chicago Linguistic Society* 7, 394–415.
- Greenberg J H (1957). 'Order of affixing: a study in general linguistics.' In *Essays in linguistics*. Chicago: University of Chicago Press.
- Greenberg J H (1978). 'How does a language acquire gender markers?' In Greenberg J H, Ferguson C A & Moravcsik E A (eds.) *Universals of human language* 2. Palo Alto, CA: Stanford University Press. 47–82.
- Hall C (1988). 'Integrating diachronic and processing principles in explaining the suffixing preference.' In Hawkins (ed.). 321–349.

- Harris A C & Campbell L (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Harris M B (1978). *The evolution of French syntax: a comparative approach*. London: Longman.
- Hawkins J A (ed.) (1988). *Explaining language universals*. Oxford: Blackwell.
- Hawkins J A & Cutler A (1988). 'Psycholinguistic factors in morphological asymmetry.' In Hawkins (ed.), 280–317.
- Hawkins J A & Gilligan G (1988). 'Prefixing and suffixing universals in relation to basic word order.' *Lingua* 74, 219–259.
- Hintze F (1947). 'Die Haupttendenzen der ägyptischen Sprachentwicklung.' *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 1, 85–107.
- Hodge C T (1970). 'The linguistic cycle.' *The Language Sciences* 13, 1–7.
- Kahr J C (1976). 'The renewal of case morphology: sources and constraints.' *Working Papers on Language Universals* 20, 107–151.
- Kraehenmann A & Plank F (forthcoming). 'Words without vowels in Upper German: clitics wanting to be affixes.' Manuscript, University of Konstanz.
- Lord C (1993). *Historical change in serial verb constructions*. Amsterdam: John Benjamins.
- Martin J B (1994). 'Implications of plural reduplication, infixation, and subtraction for Muskogean subgrouping.' *Anthropological Linguistics* 36, 27–55.
- Martin J B & Munro P (2005). 'Proto-Muskogean morphology.' In Scancarelli J & Hardy H (eds.) *The native languages of the southeastern United States*. Lincoln: University of Nebraska Press.
- Oinas F (1961). *The development of some postpositional cases in Balto-Finnic languages*. Helsinki: Suomalais-Ugrilainen Seura.
- Reh M (1986). 'Where have all the case prefixes gone?' *Afrikanistische Arbeitspapiere* 5, 121–134.
- Sapir E (1921). *Language*. New York: Harcourt, Brace, and World.
- Siewierska A & Bakker D (1996). 'The distribution of subject and object agreement and word order type.' *Studies in Language* 20, 115–161.
- Steele S (1977). 'Clisis and diachrony.' In Li C N (ed.) *Mechanisms of syntactic change*. Austin: University of Texas Press. 539–579.
- Yu A C (2003). *The morphology and phonology of infixation*. Unpublished dissertation, University of California at Berkeley.

Dinka

G J Dimmendaal, University of Cologne, Cologne, Germany

© 2006 Elsevier Ltd. All rights reserved.

Dinka is spoken by around 1.4 million people, mainly in the southern Sudan. Together with Nuer and Atuot, it forms a subgroup within Western Nilotic, one of the three primary branches of Nilotic (Nilo-Saharan). Dinka divides into four major dialects or regional variants: Padang (Northeastern Dinka), Bor (Southeastern Dinka), Rek (Southwestern Dinka), and a south-central variant known as Agar (South Central Dinka); in addition, there is a more deviant northwestern variant known as Ruweng.

Although Dinka has been studied for over 180 years, it was not until more recently, mainly as a result of investigations by the Danish scholar Torben Andersen, that important structural features of the language came to be better understood. As argued by Andersen (1987; 1990; 1993), the (Agar) Dinka vowel system involves seven vowel qualities: /i, e, ε, a, ɔ, o, u/, whereby each vowel can be either breathy or creaky; historically, the breathy/creaky distinction goes back to a distinction between [+advanced tongue root] and [-advanced tongue root] vowels (Andersen,

1990). According to Andersen (1987), there is a ternary vowel-length distinction (short vs. mid vs. long), at least in the Agar dialect of Dinka (see **Table 1**). In an alternative analysis of the same phenomena, Gilley and Remijsen (forthcoming) interpret this ternary length distinction in Rek (Southwestern) Dinka in terms of the interaction between a binary length distinction and a complementary quantity contrast. The latter feature is supposed to account for the observed covariation between centralization of vowels, nucleus and coda duration, and the realization of coda consonants in a more natural way.

Table 1 Two Dinka nouns, transcribed for segmental phonemes and quantity both in the singular and in the plural, in terms of Andersen's (1987; 1993) ternary vowel-length hypothesis, and in terms of the complementary quantity hypothesis advanced by Remijsen and Gilley (forthcoming)

| | | Ternary vowel-length hypothesis | | Complementary quantity hypothesis | |
|---------|-----|---------------------------------|-------|-----------------------------------|--------|
| 'hand' | sg. | ciin | /VV/ | cin | /VC/ |
| | pl. | cin | /V/ | cinn | /VCC/ |
| 'grass' | sg. | nooon | /VVV/ | noon | /VVC/ |
| | pl. | noon | /VV/ | noonn | /VVCC/ |

Dinka distinguishes between high and low tones, which may also be combined to form a falling tone. It is an essentially monosyllabic language that nevertheless uses various layers in its derivational and inflectional morphology, involving segmental changes in the nucleus and the coda as well as tonal modifications (compare, for example, Andersen, 1993). This internal morphology corresponds to suffixation processes in more conservative Western Nilotic languages.

Whereas traditionally Dinka has been claimed to be a SVO language, Andersen (1991) has shown that the basic position of subjects in at least one variety of Dinka, Agar, is postverbally. Postverbal (but not preverbal) subjects are marked with Nominative case by way of tonal inflection, a feature also found in other Western Nilotic languages like Anywa (Anuak), Jur Lwo (Luwo), Pāri, or Shilluk. Preverbal noun phrases are topics, whose underlying grammatical relation can be that of subject, object, adverbial, or possessor. Compare:

- (1) b̩ɔŋ ʔ-tóoc d̩ɔŋk
chief:ABS D-send boy
'the chief is sending the boy'
- (2) d̩ɔŋk ʔ-tóoc b̩ɔŋ
boy D-send:NTS chief:OBL
'the chief is sending the boy'

As further shown by Andersen (1991), such constructions are formally distinct from passives in Agar Dinka:

- (3) d̩ɔŋk ʔ-tóoc (n)ɛ b̩ɔŋ
boy D-send:PASS PREP chief:OBL
'the boy is being sent by the chief'

When no topic is expressed, sentences may be verb-initial in Agar Dinka. It remains to be determined to what extent the structure of Agar Dinka is characteristic for the Dinka cluster as a whole.

Bibliography

- Andersen T (1987). 'The phonemic system of Agar Dinka.' *Journal of African Languages and Linguistics* 9, 1–27.
- Andersen T (1990). 'Vowel length in Western Nilotic languages.' *Acta Linguistica Hafniensia* 22, 5–26.
- Andersen T (1991). 'Subject and topic in Dinka.' *Studies in Language* 15, 265–294.
- Andersen T (1993). 'Vowel quality alternation in Dinka verb inflection.' *Phonology* 10, 1–42.
- Gilley L & Remijsen B (forthcoming). 'Complementary quantity and vowel length in Dinka Rek.' *Journal of Phonetics*.

Dogon

J Bendor-Samuel, Summer Institute of Linguistics, High Wycombe, UK

© 2006 Elsevier Ltd. All rights reserved.

Dogon is spoken by about 500 000 people in north-east Mali, east of Mopti and the Niger river up to and astride the Burkina Faso border. In the past, within the Niger-Congo phylum, Dogon has been regarded as belonging to the Gur subfamily, but since the lexical and grammatical evidence for this is weak, it is now treated as an isolate within Volta-Congo. Though the Dogon people recognize themselves as one ethnic group, there are six major dialects and several additional smaller ones. The major dialects are: T̩mbo S̩ɔ, D̩nn̩ S̩ɔ, T̩r̩ S̩ɔ, Jamsay, Togo Kan and Tom̩ Kan (Bendor-Samuel *et al.*, 1989).

Dogon has a seven-vowel system and displays a limited vowel harmony system with only one vowel of the pairs *e/ɛ* and *o/ɔ* occurring in any stem. This contrast is neutralized in nasalized forms. Nasalized vowels always occur after a nasal consonant but may also occur after oral consonants and word-initially.

There is contrast between high and low tones, but this is probably best analyzed as tonal accent. Once the position of the accented syllable and its associated tone are known, the pitch of the other syllables is predictable.

Close kinship terms exhibit a very restricted vestigial noun class system. The verbal system distinguishes perfective and imperfective forms and utilizes six verbal extensions that focus on the time and kind of action. The pronominal system comprises one basic set (nominal) from which three other sets, possessive, accusative, and embedded/addressee, are derived.

The basic word order is S.O.V. Sentences frequently concatenate verbs in verb strings with subject-verb agreement marked on the sentence final verb form. Subordinating conjunctions occur clause final; other conjunctions are clause initial. Questions are marked sentence final.

There is a topic-comment construction in which the subject or object is forefronted and replaced by a pronoun in the comment. In general, a participant is introduced into a story indefinitely (a man) then

definitely (the man) and then as a pronoun (he). After the first pronominal reference, there may be zero reference.

Bibliography

Bendor-Samuel J, Olsen E J & White A R (1989). 'Dogon.' In Bendor-Samuel J (ed.) *The Niger-Congo languages: a classification and description of Africa's largest*

language family. Lanham, NY: London University Press of America. 169–177.

Bertho J (1953). 'La place des dialectes dogon de la falaise de Bandiagara parmi les autres groupes linguistiques de la zone soudanaise.' *Bulletin de l'IFAN* 15, 405–441.

Calame-Griaule G (1978). 'Le dogon.' In Barreteau D (ed.) *Inventaire des études linguistiques sur les pays d'Afrique noire d'expression française et sur Madagascar*. Paris: Conseil International de la Langue Française. 63–69.

Domari

Y Matras, University of Manchester, Manchester, UK

© 2006 Elsevier Ltd. All rights reserved.

Definitions

Domari is the language of populations that were traditionally commercial nomads (metalworkers and entertainers) throughout the Middle East and neighboring regions. Fragmented documentation exists from Azerbaijan in the north, through to Sudan in the south. There are still Domari-speaking communities in Lebanon, Syria, and Jordan; the number of speakers is unknown. The only well-documented variety is that of Jerusalem, now spoken by only up to 100 elderly people. They refer to their language as Dōmari or Dōmi. Names in other regions include Domani and Qurbati.

History

Like Romani, Domari shares a number of ancient isoglosses with the Central branch of Indo-Aryan, most notably the realization of Old Indo-Aryan *ṛ* as *u* or *i* (Sanskrit *śṛṇ-*, Domari *sun-/sin-* 'to hear') and of *kṣ-* as *k(h)* (Sanskrit *akṣi*, Domari *aki* 'eye'). It also preserves a number of clusters that have been lost in the other Central languages (Sanskrit *drākṣa*, Domari *drak* 'grape'; Sanskrit *oṣṭha*, Domari *ošt* 'lip'; Sanskrit *hasta*, Domari *xast* 'hand'). It appears therefore that Domari, like Romani, emerged as one of the Central Indic languages, but migrated prior to the loss of these clusters to the northwest, where the clusters were generally retained. Both Romani and Domari also share the pattern of renewal of the past-tense conjugation (through affixation of oblique enclitic pronouns to the past participle) with northwestern Indian frontier languages such as Kashmiri and Shina. The morphology of the two languages is similar in

other respects: both retain the old present conjugation in the verb (Domari *kar-ami* 'I do'), and consonantal endings of the oblique nominal case (Domari *mans-as* 'man.OBL', *mans-an* 'men.OBL'), and both show agglutination of secondary (Layer II) case endings (Domari *mans-as-ka* 'for the man').

It had therefore been assumed that Romani and Domari derived from the same ancestor idiom, and split only after leaving the Indian subcontinent. However, some isoglosses separating the two languages in phonology, morphology, and lexicon appear to be rather old, and point instead to a similar phenomenon of gradual northward and westward migrations, perhaps even to convergent development, rather than to a shared origin. Typical phonological developments that characterize Domari are loss of aspiration in *bh*, *dh*, *gh* to *b*, *d*, *g*; shift of medial *d*, *t* to *r*, of initial *v* to *w*, and of the retroflexes *ḍ*, *ṭ*, *ḍḍ*, *ṭṭ*, *ḍh*, etc., to *r*, *t*, and *d*.

The sound system

There is much volatility and variation in the Domari sound system. Consonants include the stops *b*, *d*, and *g*, and *p*, *t*, *k*, *q*, and *ʔ*, the fricatives *f*, *x*, *χ*, *ɣ*, and *h*, liquids *r* and *l* (and, marginally, a velarized *l*), the glides *y* and *w* (alternating with *v*), and the sibilants *s*, *z*, and *š*. The affricates *dž* and *č* alternate with their sibilant counterparts *ž* and *š*. Pharyngealization of the dentals *d*, *t*, *s*, and *z* enters the language in Arabic loans, but is often imported into the pre-Arabic (Indic) lexicon as well (e.g. /wa:tʰ/ 'stone'). The pharyngeals *ħ* and *ʕ* appear only in Arabic loans. Consonant gemination is distinctive.

Domari vowel phonemes are *a*, *e*, *i*, *o*, *ɔ*, and *u*, each showing a number of allophonic variants. Vowel length is generally distinctive, though the duration of a vowel in a given word may vary considerably. Stress normally falls on the final inflectional segment

of the word. Unstressed affixes are agglutinative (Layer II) case endings, external tense markers, and enclitic object pronouns.

Morphology

Nominal forms

The principal inflectional alternation in the noun is between two 'basic' or Layer I cases, nominative and oblique. Vocalic stems in the inherited (Indic) lexical component have the nominative endings *-a* (masculine) and *-i* (feminine). The most common oblique endings in the singular are *-as-* (*-s-* with vocalic stems) for masculines and *-ya-* (*-ē-* with vocalic stems) for feminines. Some consonantal stems, especially Arabic loans, take *-ī-* or *-ē-*. The oblique plural ending is generally *-(y)an-*. The oblique stem serves as the case of the direct object, and as the base for further (Layer II) agglutinative case formation, with the endings *-ta* (dative), *-ma* (locative), *-ka* (directive and benefactive), *-ki* (ablative and prepositional), and *-san(ni)* (instrumental and comitative).

Demonstratives and adjectives in attributive position agree with the head noun in gender, number, and case. Enclitic pronouns are used with nouns as possessive endings. They encode case and number (*putr-o-m* 'my son', *putr-i-m-ka* 'for my son', *putr-e-m* 'my sons'), as well as person (*putr-o-man* 'our son'). These enclitic pronouns also serve as object-concord markers with verbs, and as subject-concord markers with past-tense (perfective) verbs (*laked-om-ir* 'I saw you', *laked-or-im* 'you saw me'). The genitive-possessive construction marks the head with a possessive affix and the dependent in the ablative/prepositional case (*kury-os mans-as-ki* house-3SG man-OBL-ABL 'the man's house').

Verbs

Domari retains the Old Indo-Aryan intransitive (passive) derivation marker *-y-* (*ban-ari* 'shuts', *ban-y-ari* 'is being shut'). The transitive/causative marker *-naw-* is also productive (*q-ari* 'eats', *q-naw-ari* 'feeds'). The verb root with derivational augmentation constitutes the present or nonperfective stem. The perfective stem is formed by means of a perfective extension marker (*ban-ami* 'I close', *ban-d-om* 'I closed'). Arabic verb roots are integrated by means of the 'carrier' verbs *-k(ar)-* (transitive, from 'to do'), and *-h(r)-* (intransitive, from 'to become').

There are two sets of person markers. The present stem conjugation is (mostly) a direct continuation of the Old Indo-Aryan set of person markers (1SG *-m-*, 2SG *-k-*, 3SG *-r-*, 1PL *-n-*, 2PL *-s-*, 3PL *-n(d)-*). The perfective set derives partly from possessive markers in the singular, and from a combination of sources in the plural (1SG *-m-*, 2SG *-r-*, 3SG *-s-* [or M *-a*, F *-i*], 1PL *-n-*, 2PL *-s-*, 3PL *-e-*). The 3SG distinguishes between plain subjects, which show gender agreement (*kard-a* 'he did', *kard-i* 'she did'), and agentive subjects (*kard-os-is* 'he/she did it').

Tenses draw on the two stems, present and perfective, and the affixes *-i-* (progressive) and *-a-* (remote), which are external to the person affixes. The present stem followed by *-i-* constitutes the present/future tense (*laha-m-r-i* 'I see you'); followed by *-a-* it indicates the imperfect/habitual (*laha-m-r-a* 'I used to see you/was seeing you'). The perfective stem forms the basis for the simple past (*lake-d-om-ir* 'I saw you'), the perfect (*lake-d-om-r-i* 'I have seen you'), and the pluperfect/counterfactual (*lake-d-om-r-a* 'I had seen you/would have seen you').

The copula is enclitic. In the third person, predicate nouns and adjectives take a predicative suffix (M *-ēk*, F *-ik*, PL *-ēni*). Most of the modal verbs are borrowed from Arabic, and carry Arabic person and tense inflection.

Syntax

Domari shows syntactic convergence with Arabic. Word order is VO-based and flexible, and clauses are finite. All conjunctions and particles and most adverbs and numerals are borrowed from Arabic, as are most of the prepositions. While demonstratives precede the noun, there is a tendency to use adjectives mainly in predicative constructions, which agrees with the Arabic word order noun-adjective.

Bibliography

- Macalister R A S (1914). *The language of the Nawar or Zutt, the nomad smiths of Palestine*. Edinburgh: Edinburgh University Press.
- Matras Y (1999). 'The state of present-day Domari in Jerusalem.' *Mediterranean Language Review* 11, 1–58.
- Matras Y (2002). *Romani: a linguistic introduction*. Cambridge: Cambridge University Press.

Dravidian Languages

P S Subrahmanyam, Annamalai University,
Bangalore, India

© 2006 Elsevier Ltd. All rights reserved.

Languages and Subgroups

The Dravidian language family, first recognized as separate by Francis Whyte Ellis in 1816, is the fifth largest language family in the world. It consists of four widely spoken literary languages and approximately 20 minority languages (the number increases if some dialects are counted as separate languages). They are concentrated mainly in the four southern states (Tamilnadu, Kerala, Karnataka, and Andhra Pradesh) of India. Some other states, namely Maharashtra, Madhya Pradesh, Orissa, and Bihar, also have some of the tribal languages of the family, but most conspicuous is the presence of Brahui in Pakistan from ancient times. The languages are usually divided into three main groups—South Dravidian, Central Dravidian (with two subsubgroups, Telugu-Kuvi and Kolami-Parji), and North Dravidian—as shown in Table 1 (some advocate a closer relationship of Telugu-Kuvi with South Dravidian rather than with

Kolami-Parji). Of these, only four languages, Tamil, Malayalam, Kannaḍa, and Telugu, have their own scripts and literature from ancient times. The Sangam literature of Tamil, composed between the 2nd century B.C. and the 5th century A.D., and the pre-Christian-era grammar Tolkāppiyam of the same language are the oldest literary monuments for the entire family. The first inscription of Kannaḍa belongs to the 5th century A.D. and that of Telugu to the 6th century A.D.; the literature of these two languages starts in the 9th and the 11th centuries A.D., respectively. Malayalam evolved as a separate language from Old Tamil around the 9th century.

Phonology

Vowel System

Like Proto-Dravidian, most of the languages have a 10-vowel system with five short and five long ones (see Table 2) but Irūḷa, Toda, Kurumba, and Koḍagu have added centralized vowels to the inherited stock. Tuḷu developed a contrast between [e] (< [*-ēn]) and e in word-final position, for example, *kal-t-E* ‘I learned’ versus *kal-t-e* ‘he learned’. In Modern Telugu

Table 1 Dravidian language groups

| Subgroup | Language | Abbreviation | Locations | Number of speakers |
|---|-----------|--------------|--|--------------------|
| South Dravidian | Tamil | Ta. | Tamilnadu, Sri Lanka, South Africa, Malaysia, Singapore, Maritius, Fiji, Burma | 58 million |
| | Malayalam | Ma. | Kerala | 30 million |
| | Irula | Ir. | Tamilnadu | 5200 |
| | Kodagu | Kod. | Karnataka | 93 000 |
| | Kota | Ko. | Nilgiris, Tamilnadu | 1400 |
| | Toda | To. | Nilgiris, Tamilnadu | 1600 |
| | Kannada | Ka. | Karnataka, Badaga dialect in the Nilgiris, Tamilnadu | 33 million |
| | Kurumba | Ku. | Nilgiris, Tamilnadu | 5000 |
| | Tulu | Tu. | Koraga dialect, Karnataka (Koraga dialect) | 1.6 million |
| | Telugu | Te. | Andhra Pradesh | 66 million |
| Central Dravidian: Telugu-Kuvi subsubgroup | Gondi | Go. | Maharashtra, Madhya Pradesh, Orissa, Andhra Pradesh; Koya dialect in Andhra Pradesh and Orissa | 2.4 million |
| | Konda | | Andhra Pradesh and Orissa | 17 864 |
| | Pengo | Pe. | Orissa | 1300 |
| | Manda | | Orissa | Not known |
| | Kui | | Orissa | 641 662 |
| | Kuvi | | Orissa and Andhra Pradesh | 246 513 |
| | Kolami | Kol. | Maharashtra and Andhra Pradesh | 99 281 |
| Central Dravidian: Kolami-Parji subsubgroup | Naikri | | Maharashtra | 1500 |
| | Naiki | Nk. | Maharashtra | 54 000 |
| | Gadaba | Ga. | Kopdekor dialect [Andhra Pradesh], Ollari dialect [Orissa] | ?18 000 |
| | Parji | Pa. | Chattisghadh, Dhurwa | 44 000 |
| | Kurux | Kur. | Bihar | 1.4 million |
| North Dravidian | Malto | Malt. | Rajmahal hills of Bihar | 108 148 |
| | Brahui | Br. | Baluchistan of Pakistan | 1.7 million |

Table 2 Vowels of Proto-Dravidian

| | Front | | Central | | Back | |
|------|-------|------|---------|------|-------|------|
| | Short | Long | Short | Long | Short | Long |
| High | i | ī | | | u | ū |
| Mid | e | ē | | | o | ō |
| Low | | | a | ā | | |

[ɛ] (< [*iyā]), which is the past-tense suffix, appearing after most of the verb bases before personal suffixes other than 3rd nonmasculine, became a separate phoneme. Brahui lost short [e] and short [o] under the influence of neighboring Balochi. A long vowel of the root syllable undergoes shortening when the root is followed by a root extension that begins with a vowel; this alternation is regular in verb bases, less common in disyllabic nominal bases, and totally absent in trisyllabic nominal bases, for example, [*vīl-] ~ [*vīl-u-] ‘to fall’ (Burrow and Emeneau, 1984: 5430), [*nīl-al] ~ [*nīl-al] ‘shade’ (Burrow and Emeneau, 1984: 3679).

Consonant System

The consonant system as normally reconstructed is given in Table 3. The reconstruction of a laryngeal sound for Proto-Dravidian on the basis of the Old Tamil sound called *āytam*, preserved in a few words, involves speculation (Krishnamurti, 1997, 2003: 91). The most conspicuous features of the consonant system are:

1. The presence of retroflex consonants (stop, nasal, lateral, and approximant, the last one being the most peculiar feature of the entire phonological system), which are rare in the languages outside the Indian subcontinent (even the presence of these in Sanskrit and other Indo-Aryan languages is attributed to the influence of Dravidian).
2. The presence of six stops, the most peculiar of which is the alveolar (with the uncommon three-way contrast among the dental, retroflex, and alveolar ones still retained in Malayalam, Iruḷa, Kota, Toda, and Kurumba).
3. The absence of voiced stops and aspirated stops.

The initial stops in some words irregularly became voiced in all languages except Tamil-Malayalam and Toda; the voicing of single stops in medial position seems to have developed in the later stages of Proto-Dravidian itself. The apicals, that is, the alveolar and the retroflex consonants, do not occur at the beginning of a word; but this situation changed in Malayalam, Koḍagu, Tuḷu, and Telugu-Kuvi. Clusters involving two different stops do not occur within

Table 3 Consonants of Proto-Dravidian^a

| | L | D | A | R | P | Vel |
|-------------|---|---|---|---|---|-----|
| Stop | p | t | ṭ | ṭ | c | k |
| Nasal | m | n | ṇ | ṇ | ṇ | (ŋ) |
| Lateral | | | l | ḷ | | |
| Trill | | | r | | | |
| Approximant | | | | ɻ | | |
| Semivowel | v | | | | y | |

^aAbbreviations: A, alveolar; D, dental; L, labial; P, palatal; R, retroflex; Vel, velar.

a morpheme. However, homorganic clusters of a nasal (N) and plosive (P) of the types NP, PP, and NPP can be reconstructed, for example, [*marunt(u)] ‘medicine’ (Burrow and Emeneau, 1984: 4719), [*cupp(u)] ‘salt’ (Burrow and Emeneau, 1984: 2674), [*kalaṅkkam] ‘turbidity, confusion’ (Burrow and Emeneau, 1984: 1303).

Syntax

Word Classes

The following word classes can be recognized for Dravidian: nouns (pronouns and numerals are subclasses of nouns because they are inflected for case and can occur as the head of a noun phrase like nouns), verbs, adjectives, adverbs (including expressives), particles, and interjections. The first two, which are the major classes, are dealt with in detail later; a few words on each of the other categories are in order here.

Adjectives occur before nouns, for example,

- (1a) Ta.: nalla paiyaṇ
 (1b) Te.: manci abbāyi
 ‘good boy’

Some nouns are converted into adjectives by the addition of the past adjectival participle of the verb [*āk(u)-] ‘become’, for example,

- (2a) Ta.: aḷak-āna peṇ
 beauty-ADJ girl
 (2b) Te.: andam-ayina pilla
 beauty-ADJ girl
 ‘beautiful girl’

Monomorphemic adverbs are few in number; most of them are formed from nouns or adjectives by the addition of the suffix, for example, Ta. -āka, Ma. -āyi, Ka. -āgi, Te. -gā:

- (3a) Ta.: aḷak-āka
 beauty-ADV
 (3b) Te.: andaṅ-gā
 beauty-ADV
 ‘beautifully’

The following are examples of expressives:

- (4a) Ma.: avan karumure ti-nnu
he karumure eat-PAST
'he ate with a crunching sound'
- (4b) Ka.: avanu paṭapaṭane hoḍe-d-a
he paṭapaṭa beat-PAST-3.SING
'he beat (someone) thoroughly'
- (4c) Te.: vāḍu karakarā namil-Ē-ḍu
he karakara munch-PAST-3.MASC.SING
'he munched with a crunching sound'

Particles are bound forms that can be added to a wide range of major sentence constituents. Examples are:

1. The interrogative particle, Ta. Ka. Te. *-ā*, Ma. *-ō*:

- (5a) Ta.: avan va-nt-ān-ā?
he come-PAST-3.MASC.SING-INT
- (5b) Ma.: avan va-nn-ō?
he come-PAST-INT
'did he come?'

2. The particle of emphasis, Ta. *-tān*, Ma. *-tanne*, Ka. Te. *-ē*:

- (6a) Ta.: avan tān
he-EMPH
- (6b) Te.: vāḍ ē
he-EMPH
'he himself'

3. The particle of coordination, added at the end of both (or all) the coordinated noun phrases, Ta. Ma. *-um*, Ka. *-ū*, Te. lengthening of the final vowel (but *-ū/-Ø* after nouns ending in [m]):

- (7a) Ta.: pāl-um paḷam-um
milk-COORD fruit-COORD
'milk and fruit'
- (7b) Te.: mogudū peḷḷām(-ū)
husband-COORD wife-COORD
'husband and wife'

Examples for interjections are: Ta. *ām* (spoken, [āmā]), Ma. *ate*, Ka. *haudu*, Te. *av(u)nu* 'yes'; Ta. Ma. Ka. Te. *ayyō/ayyayyō* expression of surprise, sympathy, pain, grief, or fear, *chī* expression of disgust.

Articles, conjunctions, and dummy subjects (as *it* in *it is raining*) are absent in Dravidian.

Word Order

The unmarked sentence structure is S(ubject) O(bject) V(erb). The head of the subject noun phrase is in the nominative case, whereas that of the object noun phrase is in the accusative, the suffix for which can be unmarked in the case of inanimate nouns (see later discussion). Other noun phrases that can be in a sentence include those that indicate the indirect object and the person associated with the agent, time and

place. Although the verb phrase normally occurs at the end of the sentence, the noun phrases can exchange their positions within the sentence with some freedom, as illustrated by the Telugu sentences in (8).

- (8a) rāmuḍu sīta-ni ninna
Rama Sita-ACCUS yesterday
poddunna cūs-Ē-ḍu
morning see-PAST-3.MASC.SING
'Rama saw Sita yesterday morning.'

which can also appear as:

- (8b) rāmuḍu ninna poddunna sīta-ni cūs-Ē-ḍu
(8c) sīta-ni rāmuḍu ninna poddunna cūs-Ē-ḍu
(8c) ninna poddunna rāmuḍu sīta-ni cūs-Ē-ḍu
(8d) ninna poddunna sīta-ni rāmuḍu cūs-Ē-ḍu

and so on.

Agreement

When used as the predicate, a noun shows agreement in gender and number with the subject 3rd-person pronoun and, when used as the subject, it shows agreement with the finite verb (for the latter, Malayalam is an exception). Therefore, nouns are classified on the basis of their gender and number, but there is no uniformity among the languages in this matter. Toda and Brahui have no gender distinction at all; original nonhuman forms have been generalized in these languages for all categories, for example, To. *aθ*, Br. *ōd* 'he, she, it'. The other languages can be classified into four groups on the basis of the gender-number distinctions they show. The South Dravidian languages show a five-way distinction, as in Table 4. Telugu and Kuṛux-Malto show a four-way distinction, as in Table 5. (Te. *āmelāviḍa* 'she (honorific)' has no verb form that exclusively corresponds to it and takes either the nonmasculine singular or the human plural form.) Pengo and Maṇḍa have a six-way contrast in the pronoun, as shown in Table 6, but the contrast between the feminine singular and the nonhuman singular is neutralized in the verb. The central languages other than Telugu, Pengo, and Maṇḍa have a symmetrical system with a four-way distinction, as shown in Table 7.

Table 4 Gender-number distinctions: Tamil

| | Human | | Nonhuman |
|----------|---------------------------|---------------|-------------------------|
| | Masculine | Feminine | |
| Singular | avan 'he' | avaḷ 'she' | atu 'it' |
| Plural | avar(kal) 'they (HUM)' | | avai 'they (NONHUM)' |

Table 5 Gender-number distinctions: Telugu

| | <i>Masculine</i> | <i>Nonmasculine</i> |
|----------|-----------------------|------------------------|
| Singular | vāḍu 'he' | adi 'she/it' |
| | <i>Human</i> | <i>Nonhuman</i> |
| Plural | vā u 'they (HUM)' | avi 'they (NONHUM)' |

Table 6 Gender-number distinctions: Pengo

| | <i>Human</i> | | <i>Nonhuman</i> |
|----------|--------------------------------------|-----------------------|-----------------------------|
| | <i>Masculine</i> | <i>Feminine</i> | |
| Singular | avan 'he' | adel 'she' | adi/adaŋ 'it' |
| Plural | avar 'those men/men and women' | avek 'those women' | avaŋ 'those (NONHUM)' |

Table 7 Gender-number distinctions: Gondi

| | <i>Masculine</i> | <i>Nonmasculine</i> |
|----------|-------------------------------------|----------------------------|
| Singular | vōr 'he' | ad 'she, it' |
| Plural | vūr 'those men/men and women' | av 'those women/things' |

Equational Sentences without the Copula

A characteristic feature of the major Dravidian languages is the presence of equational sentences without the copular verb; however, Malayalam and many of the central and the northern languages have innovated by creating the copula under the influence of Indo-Aryan.

- (9a) Ta.: en peyar kumār
my name Kumar
'my name is Kumar'

but:

- (9b) Ma.: en-re pērō kumār enn(ō) āṇō
I-GEN name Kumar say-PAST COP
'my name is Kumar'

Dative-Subject Sentences

Although not exclusive to Dravidian, sentences with a dative subject are commonly used in these languages. In sentences of this type, the logical subject (i.e., the

noun that denotes the person or animate being who has some feeling, such as anger, hunger, wanting, or liking, or has/acquires/needs something, abstract or concrete) appears in the dative case and the noun that denotes the feeling or the thing is in the nominative and serves as the surface subject.

- (10a) Ta.: en-akku kōpam va-nt-atu
I-DAT anger come-PAST-3.N.SING
'I got angry'
(10b) Te.: vāḍ-i-ki dabbu kāvāli
he-OBL-DAT money is required
'he needs money'

Complex Sentences

The most notable among the complex sentences are those (1) with a past adverbial participle, (2) with noun phrases with a relative participle (also known as verbal adjective), and (3) with the quotative marker.

A special feature of the Dravidian languages is the possibility of having more than one past participle in a sentence; this feature has spread from Dravidian to Sanskrit and modern Indo-Aryan. Examples are:

- (11a) Ka.: vanaja mane-ge hōg-i snāna māḍ-i
Vanaja house-DAT go-PAST bath do-PAST
baṭṭe badalāyis-i ūṭa māḍ-id-aḷu
clothes change-PAST food do-PAST-
3FEM.SING
(11b) Te.: vanaja iṭṭi-ki vell-i snānam cēs-i
Vanaja house-DAT go-PAST bath do-PAST
baṭṭa-lu mārcukun-i annam tin-di
cloth-PL change-PAST food eat-PAST.3.
FEM.SING
'Vanaja went home, took a bath, changed her
clothes and ate food'

The Dravidian languages do not have relative pronouns; their functions are carried out by the verbal adjectives, which are derived from verb bases and function as adjectives carrying tense distinction and even negation (see later discussion). A verbal adjective can occur before the head noun (which may be preceded by other types of adjectives).

- (12) Te.: ninna poddunna meḍrāsu-occ-ina
yesterday morning Madras- come-
ABL PAST.RP
mā tammuḍu ikkaḍa unnāḍu
our (excl) younger here be-PRES-3.
brother MASC.SING
'my younger brother who came from Madras
yesterday morning is here'

A verbal adjective can qualify not only the agent of the source verb, as in (12), but also nouns denoting other relations, such as object (13a) and instrument (13b).

- (13a) Ta.: kumār **ceyy-um** vēlai
Kumar do-NON-PAST.RP work
'the work that Kumar does'

- (13b) Te.: kumār paṇḍu **kōs-ina** katti
Kumar fruit cut-PAST.RP knife
'the knife with which Kumar cut the fruit'

A quoted sentence precedes the matrix clause, which contains such verbs as 'tell', 'say', 'think', 'ask', 'hear', 'believe', and 'know'. The quotative marker (e.g., Ta. *enru*, Ma. *ennō*, Ka. *endu*, Te *ani* 'having said', past participles of the verb 'to say') occurs at the end of the quoted sentence.

- (14) Te.: dēvuḍu unnāḍu **an-i**
god be-PRES-3MASC.SING say-PAST
cālā mandi nammu-tā-ru
many people believe-FUT-3.HUM.PL
'many people believe that there is god'

This Dravidian construction has influenced the quotative construction in Sanskrit as shown by the marker *iti*, which follows the quoted sentence.

Noun Morphology

Dravidian morphological structure is agglutinative. Nouns in Dravidian are mostly underived (e.g., Ta. Ma. Te. *puli* 'tiger'), but there are also nouns derived from verbs (e.g., Te. *pāṭa* 'song' from *pāḍu-* 'to sing') and from adjectives (e.g., Ta. *nalla-tu* 'good thing' from *nalla* 'good'). A nominal stem may be followed, when plurality has to be expressed, by a plural suffix, which in turn may be followed by a case suffix or a (case suffix +) postposition (i.e., a separate word with case function), as in:

- (15a) Ta.: kulantai-kaḷ-ukku
child-PL-DAT

- (15b) Te.: pilla-la-ki
child-PL-DAT
'to the children'

Only a few of the nominal stems contain an overt marker for gender. For example, in Tamil, *aṇṇa-n* 'elder brother' has the masculine suffix but *tampi* 'younger brother' has no suffix.

Plural Suffixes

The plural suffix in most of the languages shows a distinction for human vs. nonhuman (or for masculine vs. nonmasculine in the Central group of languages other than Telugu) with some exceptions in which the nonhuman suffix occurs with human nouns, as in Kannaḍa:

- (16a) huḍuga-ru
boy-PL.HUM
'boys'

- (16b) mara-ḡaḷu
tree-PL.NHUM
'trees'

but:

- (16c) mantri-ḡaḷu
mnister-PL.NHUM
'ministers'

There are also some languages in which the erstwhile nonhuman plural suffix is generalized to all nouns at the expense of the human plural suffix, as in Telugu:

- (17a) akka-lu
'elder sisters'

- (17b) nakka-lu
'jackals'

The Toda plural suffix *-ām* (optional), traceable to the postnominal modifier [**anayttum*] 'all (NONH.)' illustrates the process of a separate word being reduced to the status of a suffix over course of time, as in To. *kaṣ-ām* 'stones' (cf. Ta. *kall anaittum* 'all the stones'). The use of the plural suffix with nonhuman nouns is rare in the Southern and the Northern groups, whereas it is obligatory in the Central group, for example,

- (18a) Ta.: iraṇṭu paḷam
'two fruit'

but:

- (18b) Te.: reṇḍu paḷlu (< *paṇḍu-lu)
'two fruit'

([*reṇḍu paṇḍu] without the plural *-lu* is ungrammatical in Telugu.)

Case Suffixes

The nominative is unmarked in all the languages. Some of the nominal stems take an additional suffix, called the oblique suffix, when a case suffix other than the nominative or a postposition is added; the stem thus formed is called the oblique stem. For example, in Tamil, *vītu* 'house' and *maram* 'tree' have the oblique stems *vīt-t-* (as in *vīt-t-il* 'in the house') and *mara-tt-* (*mara-tt-il* 'in the tree'), respectively, but others, such as *ūr* 'village' (*ūr-il* 'in the village') and *pāl* 'milk' (*pāl-il* 'in the milk') do not have a separate oblique stem.

With regard to the accusative suffix, the languages are divided into two main groups with Tamil, Malayalam, Koḍagu, Irūḷa, Kurumba, and Brahui showing the reflexes of [**-ay*], whereas all others show the reflexes of [**-n*], often preceded or followed by a vowel.

- (19a) Ta.: nāy-ai
dog-ACCUS

- (19b) Ma.: paṭṭiy-e
dog-ACCUS
(19c) Ka.: nāy-annu
dog-ACCUS
(19d) Te.: kukka-ni
dog-ACCUS
'dog'

The two were probably dialectal variants in Proto-Dravidian. The accusative is generally unmarked with inanimate nouns in all the languages, for example,

- (20a) Ta.: nān pāl kuṭi-tt-ēn
I milk drink-PAST-1.SING
(20b) Te.: nēnu pālu tāg-Ē-nu
I milk drink-PAST-1.SING
'I drank milk'

In Gonḍi (except the Koya dialect), Brahui, and some other non-South Dravidian languages, the dative case also assumes the functions of the accusative under the influence of Indo-Aryan:

- (21) Go.: vōr nā-kūn sūr-t-ōr
he I-ACCUS/DAT see-PAST-3.MASC.SING
'he saw me.'

The instrumental case is also used in the sociative sense in many languages, such as Telugu:

- (22a) nēnu karra-tō kukka-ni koṭṭ-Ē-nu
I stick- dog-ACC beat-PAST-3.
INSTR/SOC 1.SING
'I hit the dog with a stick'
(22b) nēnu ayana-tō māṭṭāḍ-Ē-nu
I he-INSTR/SOC speak-PAST-1.SING
'I spoke with him'

But the two are distinguished in Tamil and some others:

- (23a) Ta.: kaiy-āl
hand-INSTR
'with the hand'
(23b) Ta.: avan-ōṭu/oṭu-utan
he-SOC
'with him'

Of all the case suffixes, the dative case suffix is found with minimum variation in all the subgroups and is reconstructable as [*-kk(u)] for Proto-Dravidian.

Some languages have a genuine suffix for the ablative case, whereas others form it on the locative by the addition of a postposition, which originally was the past participle of the verb 'to be' (Ta. *iru-ntu*) or 'to stand' (Ma. *ninnḍ*). Both types are illustrated by Tamil. Although Old Tamil contains the suffix *-in*, Modern Tamil has locative *-il-iruntu*, as in

- (24a) Old Ta.: mala¹iy-in
hill-ABL
'from the hill'

- (24b) Mod. Ta.: mala¹iy-il-iruntu
hill-LOC-PO
'from the hill'
(24c) Ma.: vīṭ-ṭ-il-ninnḍ
house-OBL-LOC-PO
'from the house'

The sense of the genitive case can be expressed just by the juxtaposition of one noun (or its oblique base form, if there is one) and another noun (as in the Tamil and Telugu examples in (25a) and (25b)), but a suffix or postposition is also found (as in the Kannaḍa and Koḍagu examples (25c) and (25d)).

- (25a) Ta.: tāy ccol
mother word
'mother's word'
(25b) Te.: amma māṭa
mother word
'mother's word'
(25c) Ka.: avar-a mane
they-GEN house
'their house'
(25d) Koḍ.: aynḡa-ḍa mane
they-GEN house
'their house'

What appear to be the suffixes of the locative case in the major languages are in reality postpositions (historically, in some cases). Thus, Ta. Ma. *-il* is from [*il] 'house' (Burrow and Emeneau, 1984: 494) and Ka. *-alli* is identical with *alli* 'there':

- (26a) Ta.: vīṭ-ṭ-il
house-OBL-LOC
(26b) Ka.: maney-alli
house-LOC
'in the house'

Pronouns

The following peculiarities may be noted in the pronominal system of Dravidian:

1. The distinction between exclusive (i.e., excluding the hearer) and inclusive (i.e., including the hearer) in the first-person plural, as shown in Table 8. (Modern Kannaḍa lost this distinction and has *nāvu* in both the senses.)
2. The marking of two degrees of proximity to the speaker in the third-person pronouns (e.g., Ta.

Table 8 Exclusive and inclusive 'we'

| Language | Exclusive 'we' | Inclusive 'we' |
|----------|----------------|--------------------|
| Ta. | nāṅka! | nām |
| Ma. | ñāṅṅa! | namma! / nām / nōm |
| Te. | mēm | manam |

Table 9 Politeness in the third-person singular pronouns

| Language | Nonhonorific | | Honorific | |
|----------|----------------------|--------------------|--|--------------------------------|
| | Masculine | Feminine | Masculine/feminine | |
| Ta. | avan 'he | aval 'she | avar(kal) 'he/she (HON)' | |
| | (NHON)' Masculine | (NHON) Feminine | Masculine | Feminine |
| Te. | vāḍu | adi | atanu | āme/āviḍa/ (rarely) vāru |
| | 'he (NHON)' | 'she (NHON)' | 'he (HON 1st degree)' āyana/ (rarely) vāru | 'she (HON)' |
| | | | 'he (HON 2nd degree)' | |

avan, Te. *vāḍu* 'he (remote [formed on the root *a*-])' vs. Ta. *ivan*, Te. *viḍu* 'he (proximate [formed on *i*-])'. (Old Tamil and Old Kannaḍa, Kuvi, Kurux, and Brahui show three degrees of proximity, the third one, intermediary, formed on [*u-], whereas Kui shows four degrees, the stems being *i*- [proximate], *e*- [intermediary], *a*- [remote], and *o*- [very remote].)

3. The two-way (in Tamil and Malayalam) or three-way distinction (in Telugu [only in the masculine] and Kannaḍa) based on politeness in the third-person singular pronouns, as shown in **Table 9**.
4. The use of reflexive pronouns [**tān*] (SING) and [**tām*] (PL) to refer to the third-person subject of the sentence, as in:

(27) Te.: kumār tana ḍabbu aḍig-Ē-ḍu
Kumar REFL- money ask-PAST-
 GEN 3.MASC.SING
'Kumar asked for his (own) money'

Numerals

The cardinal numerals show a distinction between nonhuman and human at the morphological level. Only the South Dravidian languages (especially the three literary languages) and Telugu show a developed native numeral system in which the highest native numeral is Te. (Old) *vēyil*/(Mod) *veyyi* ‘1000’ (PL *vē-lu*; also (Old) *vē-vuru* ‘1000 people’) (Burrow and Emeneau, 1984: 5404); the words for ‘hundred thousand’ (Ta. *laṭcam*, Ma. *lakṣam*, Ka. Te. *lakṣa*) and for ‘10 million’ (Ta.Ma.Ka.Te. *kōṭi*) are borrowed from Sanskrit and, in South Dravidian, even the word for ‘1000’ is from that source through Prakrit (e.g., Ta. Ma. *āyiram*, Ka. *sāvira* < Skt. *śahasra*-). The

Table 10 Proto-Dravidian nonhuman numerals

| <i>Numeral</i> | <i>Proto-Dravidian</i> | <i>Comment</i> |
|----------------|------------------------|--|
| 1 | *ont̪(u) | |
| 2 | *iraṇ̪t̪(u) | |
| 3 | *mūnt̪(u) | |
| 4 | *nāl̪ | |
| 5 | *cay(-nt̪(u)) | |
| 6 | *cāt̪(u) | |
| 7 | *ēl̪ | |
| 8 | *eṇ̪ | Identical with *eṇ̪ 'number' (Burrow and Emeneau, 1984) |
| 9 | *toḷ̪ | Preserved in Ta. Ma. toḷ̪!-āyiram '900', etc. |
| 10 | *patt̪(u) | |
| 100 | *nūt̪(u) | |

languages of the Central (other than Telugu) and the Northern groups retain only a few of the basic Dravidian numerals and have borrowed the higher ones from the neighboring major languages. The basic nonhuman numerals that can be reconstructed for Proto-Dravidian are shown in **Table 10**. The human forms are derived from the nonhuman ones (or their variants, in some cases) by the addition of the suffix [**-var*] ([**-van*]) also in the case of ‘one’, for example, [**oru-van*]/(honorific) [*oru-var*] ‘one man’ and [**iru-var*] ‘two persons’.

Modern Telugu and Modern Kannaḍa add a classifier (Te. *mandi*, Ka. *mandiljana* ‘people’) to the basic numeral to form the human numeral, perhaps due to influence from the neighboring Indo-Aryan languages:

(28a) Ka.: mūru mandi/jana makkaḷu
 three CLASS children
 ‘three children’

(28b) Te.: enimiḍi mandi pillalu
 three CLASS children
 ‘eight children’

The languages of the Kolami-Parji group have created separate feminine forms for the numerals ‘two’ to ‘five’ by adding the feminine suffix [**-ā]* to the root:

(29a) *īr-āl
two-FEM
'two women'
(29b) *muy-āl
three-FEM
'three women'
(29c) *nall-āl
four-FEM
'four women'
(29d) *ceyy-āl
five-FEM
'five women'

The multiples of 10 from '20' to '90' are formed by adding [**patt(u)*] '10' to the adjectival or root form of the numerals 'two' onward: [**iru-patt(u)*] '20 (lit., two tens)', and so on. The numerals '11' to '19' and numbers between multiples of 10 are formed by adding the basic numeral to '10' or its multiple, as in Ta. *pat-in-onru*, Te. *padakonḍu* '11'.

The ordinals are formed from the nonhuman cardinals by the addition of a suffix:

- (30a) Ta.: iraṇṭ-ām/-āvatu
two-ORD
(30b) Ma.: raṇṭ-ām(atte)
two-ORD
(30c) Ka.: eraḍ-aneya/-anē
two-ORD
(30d) Old Te.: reṇḍ-agu/-ava
two-ORD
(30e) Mod. Te.: reṇḍ-ava/-ō
two-ORD
'second'

Verb Morphology

The most notable peculiarities of the Dravidian verb are (1) the presence of forms with a suffix for the negative, and (2) the verbal adjectives (or adjectival/relative participles). Verbs may be divided into two main categories: finite and nonfinite. The verb base (VB), which is the basis for verb inflection and derivation, has four types. The underlying verb base may be intransitive (e.g., Ta. *nil-* 'to stand') or inherently transitive (e.g., Ta. *cey-* 'to do'). An intransitive verb base may optionally be extended by a transitive suffix, resulting in a derived transitive. A transitive verb base of either type may be converted into a causative by the addition of a causative suffix. The Telugu verb bases *taḍiyu-* 'to become wet' (with no suffix), *taḍu-pu-* 'to make (something) wet' (with the transitive *-pu-*), *taḍi-p-iñcu-* 'cause (someone) to make wet' (with transitive *-pu-* and causative *-iñcu-*), and *cēy-iñcu* 'to cause to do' (with causative *-iñcu*) illustrate the four types of verb base.

Finite Verbs

The structure of a finite verb, which is sentence-final, may be symbolized as follows:

VB + Tense/Negative Suffix + Personal Suffix

There are three major exceptions to this statement:

1. Malayalam has lost the personal suffixes completely over course of time, although they were there in its earlier stage, for example,

- (31) *ñān/nī/avan* *va-nnu*
 come-PAST
'I/you (SING)/he came'

2. In the past negative morphological forms found in the Central languages other than Telugu and Gonḍi, the negative suffix and the past-tense suffix co-occur in that order:

- (32) Kol.: *siyy-ē-t-ēn*
 give-NEG-PAST-3.SING
'he did not give'

3. In Old Tamil, the additional past suffix *-an-* is often added after a regular past suffix other than *-in-*:

- (33a) *cey-t-ār*
 do-PAST-3.HUM.PL
(33b) *cey-t-an-ar*
 do-PAST-PAST-3.HUM.PL
'they (HUM) did'

Tenses

All three tenses, past, present, and future, are found in most of the languages, as in Tamil:

- (34a) *cey-t-ēn*
 do-PAST-1.SING
'I did'
(34b) *cey-kinr-ēn*
 do-PRES-1.SING
'I am doing'
(34c) *cey-v-ēn*
 do-FUT-1.SING
'I (will) do'

But some languages have only two, past and nonpast, as in Kannada:

- (35a) *ba-nd-e*
 come-PAST-1.SING
'I came'
(35b) *baru-tt-ēne*
 come-PRES/FUT-1.SING
'I am coming/(will) come'

The present tense in many languages is periphrastic or is derived from a periphrastic construction. The future or nonpast tense also serves as the habitual.

Negative Finite Verbs

The negative finite verb, which normally denotes negation in the future or the habitual, contains the negative suffix (which may also be zero) between the verb base and the personal suffix (there is no tense marker in this construction):

- (36a) Old Ta.: *ceyy-Ø-ēn*
 do-NEG-1.SING
'I will/do not do'
(36b) Te.: *cepp-a-nu*
 tell-NEG-1.SING
'I will/do not tell'

Negation in the past and the present is expressed by syntactic constructions involving the negative auxiliary verb.

• Past negative: VB + Infinitive + Negative Auxiliary

- (37a) Ta.: coll-a(v) **illai**
 tell-INF NEG.AUX
 (37b) Te.: cepp-a **lēdu**
 tell-INF NEG.AUX
 ‘(one) did not tell’

• Present negative: Verbal Noun + Negative Auxiliary

- (38a) Ta.: col-v-at(u) **illai**
 tell-FUT-N.SING NEG.AUX
 (38b) Te.: cepp-aḍam **lēdu**
 tell-NOM NEG.AUX
 ‘(one) is not telling’

Personal Suffixes

Personal suffixes distinguish number in the first and second persons but number and gender in the third person (see Table 11). Whereas Modern Malayalam has no personal suffixes at all, Toda and Koḍagu have no personal suffixes in the third person. Because of the presence of these suffixes in the verb, the subject pronoun or noun (the latter, if it can be retrieved from the context) can be freely omitted.

Imperative

The verb base itself serves as the imperative form in the singular in most of the languages, but a special suffix is added to it in the plural, as in Table 12. The

corresponding negative imperative has the negative suffix in between the verb base and the imperative suffix, as in Table 13.

Nonfinite Verbs

The nonfinite forms may be divided into verbal adjectives and other forms, most of which serve as heads of subordinate clauses. A verbal adjective is formed by adding to the VB the tense or the negative marker, followed by the adjective marker:

• Past verbal adjective: VB + Past + Adjective

- (39a) Ta.: cey-t-a
 do-PAST-ADJ
 (39b) Te.: cēs-in-a
 do-PAST-ADJ
 ‘that did’

• Present verbal adjective: VB + Present + Adjective

- (40a) Ta.: cey-kinr-a
 do-PRES-ADJ
 (40b) Te.: cēs-tōnn-a
 do-PRES-ADJ
 ‘that is doing’

• Future/habitual verbal adjective: VB + Fut/Hab + Adjective

- (41a) Ta.: ceyy-um
 do-FUT/HAB.ADJ
 (41b) Te.: cēs-ē
 do-FUT/HAB.ADJ
 ‘that (will) do(es)’

• Negative verbal adjective: VB + Negative + Adjective

- (42a) Ta.: ceyy-āt-a
 do-NEG-ADJ

Table 11 Past tense paradigms of Ta. cey- and Te. ce:yu- ‘to do’

| Tamil | Telugu | Gloss |
|-----------------------|------------------------|--------------------|
| cey-t-ēn | cēs-Ē-nu | ‘I did’ |
| do-PAST-1.SING | do-PAST-1.SING | |
| cey-t-ōm | cēs-Ē-m | ‘we did’ |
| do-PAST-1.PL | do-PAST-1.PL | |
| cey-t-āy | cēs-Ē-vu | ‘you (sg.) did’ |
| do-PAST-2.SING | do-PAST-2.SING | |
| cey-t-īrkaḷ | cēs-Ē-ru | ‘you (pl.) did’ |
| do-PAST-2.PL | do-PAST-2.PL | |
| cey-t-ān | cēs-Ē-ḍu | ‘he did’ |
| do-PAST-3.MASC.SING | do-PAST-3.MASC.SING | |
| cey-t-āḷ | — | ‘she did’ |
| do-PAST-3.FEM.SING | | |
| cey-t-ārkaḷ | cēs-Ē-ru | ‘they (HUM) did’ |
| do-PAST-3.HUM.PL | do-PAST-3.HUM.PL | |
| cey-t-atu | cēs-in-di | (Ta.) ‘it did’ |
| do-PAST-3.NONHUM.SING | do-PAST-3.NONMASC.SING | (Te.) ‘she/it did’ |
| cey-t-aṇ-a | cēs-Ē-yi | ‘they (NEUT) did’ |
| do-PAST-3.NHUM.PL | do-PAST-3.NHUM.PL | |

Table 12 Positive imperatives

| Language | Singular | Plural | Gloss |
|----------|----------|-------------|---------|
| Ta. | col | coll-uṅkaḷ | ‘tell!’ |
| | | tell-IMP.PL | |
| Te. | ceppu | cepp-aṇḍi | ‘tell!’ |
| | | tell-IMP.PL | |

Table 13 Negative imperatives

| Language | Singular | Plural | Gloss |
|----------|-----------------|-----------------|----------------|
| Ta. | coll-āt-ē | coll-āt-īrkaḷ | ‘do not tell!’ |
| | tell-NEG-IMP.SG | tell-NEG-IMP.PL | |
| Te. | cepp-aku | cepp-ak-aṇḍi | ‘do not tell!’ |
| | tell-NEG | tell-NEG-IMP.PL | |

- (42b) Te.: ceyy-ani
do-NEG.ADJ
'that will/do(es) not do'

• Past participle: VB + Past

- (43a) Ta.: va-ntu
come-PAST
(43b) Te.: occ-i
come-PAST
'having come'

• Present participle: VB + Present

- (44a) Ka.: bar-uttā
come-PRES
(44b) Te.: os-tū
come-PRES
'coming'

• Negative participle: VB + Negative

- (45a) Ta.: ceyy-āmal
do-NEG
(45b) Te.: ceyy-akuṇḍā
do-NEG
'without doing'

• Infinitive: VB + Infinitive

- (46a) Ta.: ceyy-a
do-INF
(46b) Ma.: ceyy-ān
do-INF
'to do'

• Conditional: VB + Past + Conditional

- (47a) Ta.: cey-t-āl
do-PAST-COND
(47b) Ka.: māḍ-id-ar-e
do-PAST-COND-ADD
'if (one) does'

• Concessive: Conditional + Coordinative particle

- (48a) Ta.: cey-t-āl-um
do-PAST-COND-COORD
(48b) Ka.: māḍ-id-ar-ū
do-PAST-COND-COORD
'even though one did/does'

• Verbal noun: (49a) VB + Future + *atu* 'it', (49b) VB + Nominalizer

- (49a) Ta.: cey-v-atu
do-FUT-N.SING
(49b) Te.: ceyy-aḍam/-aṭam
do-NOM
'doing'

Auxiliary Verbs

A modal (or passive) auxiliary follows the infinitive of the main verb, whereas other types of auxiliary

verbs follow the past participle of the main verb. Tense and personal suffixes, when they can be added, are added only to the auxiliary. Examples for a modal auxiliary are:

- (50a) Ta.: avan var-a vēṇṭum
he come-INF MUST
'he must come'
(50b) Te.: vāḍu tin-āli
he eat-MUST
'he must eat'

Ta. *viṭu-* and Te. *vēyu-* (with [v] > Ø), which literally mean 'leave' and 'throw', respectively, serve here as examples for a nonmodal auxiliary; when used as an auxiliary, they mean 'completion, quickness':

- (51a) Ta.: avan paṇam koṭu-ttu viṭu-v-ān
he money give-PP leave-FUT-3.
MASC.SING
(51b) Te.: vāḍu ḍabbu icc-ēs-tā-ḍu
he money give-(PP)-throw-FUT-
3.MASC.SING
'he will give the money completely/
quickly'

The passive voice, found only in the written (not in the spoken) varieties of the literary languages, is a syntactic construction with the auxiliary Ta. *paṭu-*, Ma. *peṭu-*, Te. *paḍu* (with the change [p] > [b] in the last language), which literally mean 'fall, suffer', following the infinitive of the main verb:

- (52a) Ta.: ceyy-a ppaṭ-t-atu
do-INF PASSAUX-PAST-3.N.SING
(52b) Te.: ceyy-a baḍ-in-di
do-INF PASSAUX-PAST-3.N.SING
'it was done'

Contact between Dravidian and Indo-Aryan

Dravidian and Sanskrit (and later Indo-Aryan) show mutual influence on a large scale, which justifies calling the Indian subcontinent a linguistic area (Emeneau, 1954, 1956, 1980). A few words and structural features of Dravidian origin that have been found in the *R̥gveda* allow us to conclude that the Dravidian languages were spoken in the northwestern part of the subcontinent at that time. Two examples of Dravidian words in the *R̥gveda* are *mayūra-* 'peacock' and *khāla-* 'threshing floor, granary'. The most important structural features that spread from Dravidian to Indo-Aryan are retroflexes, the use of the past participle, the use of the Sanskrit quotative marker *iti*, the use of Sanskrit *api* in the meanings 'even, also, and, indefinite', and expressives. The Dravidian languages, in turn, have a large number of loanwords from Sanskrit.

Bibliography

- Andronov M S (2001). *A comparative grammar of the Dravidian languages*. Moscow: The Institute of Oriental Studies (Russian Academy of Sciences).
- Burrow T (1968). *Collected papers on Dravidian linguistics*. Annamalainagar, India: Annamalai University.
- Burrow T & Emeneau M B (1984). *A Dravidian etymological dictionary* (2nd edn.). Oxford: Clarendon Press.
- Caldwell R (1913/1956). *A comparative grammar of the Dravidian or south-Indian family of languages*. (reprint edn.). Madras, India: University of Madras.
- Emeneau M B (1954). 'Linguistic prehistory of India.' *Proceedings of the American Philosophical Society* 98, 282–292. (Reprinted in Emeneau, 1980: 85–104.)
- Emeneau M B (1956). 'India as a linguistic area.' *Language* 32, 3–16. (Reprinted in Emeneau, 1980: 86–104.)
- Emeneau M B (1980). *Language and linguistic area*. Dil A S (ed.). Stanford, CA: Stanford University Press.
- Emeneau M B (1994). *Dravidian studies: selected papers*. Delhi, India: Motilal Banarsidass.

- Krishnamurti B (1997). 'Proto-Dravidian laryngeal *H revisited.' *Pondichery Journal of Dravidian Studies* 7(2), 145–165.
- Krishnamurti B (2003). *The Dravidian languages*. Cambridge, UK: Cambridge University Press.
- Shanmugam S V (1971). *Dravidian nouns: a comparative study*. Annamalainagar, India: Annamalai University.
- Steever S B (ed.) (1998). *The Dravidian languages*. London and New York: Routledge.
- Subrahmanayam P S (1971). *Dravidian verb morphology: a comparative study*. Annamalainagar, India: Annamalai University.
- Subrahmanayam P S (1983). *Dravidian comparative phonology*. Annamalainagar, India: Annamalai University.
- Zvelebil K V (1990). *Dravidian linguistics: an introduction*. Pondicherry, India: Pondicherry Institute of Linguistics and Culture.
- Zvelebil K V (2001). *Nilgiri areal studies*. Prague: Charles University (Karolinum Press).

Dutch

J M Y Simpson, University of Aberdeen,
Aberdeen, UK

© 2006 Elsevier Ltd. All rights reserved.

Dutch (Nederlands) is the official language of the Netherlands (with approximately 14 million speakers) and one of the official languages of Belgium (with approximately 6 million speakers). It is also the language of administration in Aruba, Netherlands Antilles, and Suriname. There are approximately 410 000 speakers of Dutch in the United States, 159 000 in Canada, 80 000 in France, and 47 000 in Australia. The name Flemish (Vlaams), formerly applied to all varieties of Dutch spoken in Belgium and France, is now properly used only of the dialects of the Belgian provinces of West and East Flanders (West- and Oost-Vlaanderen).

Genetic Relationship

Dutch (D), together with English (E) and Frisian, belongs to the Low German branch of the West Germanic languages. That it has not participated in the Second (or High German) Sound Shift is shown by such forms as *schaap*, *beter*, *boek* (= E *sheep*, *better*, *book*) compared to (High) German (G) *Schaf*, *besser*, *Buch*. Other features shared with English are the relative rarity of grammatical umlaut (D *boek*, *boeken*, E *book*, *books* vs. G *Buch*, *Bücher*) and

the retention of the consonant clusters [sp] and [st] (D *spreken*, *stelen*, E *speak*, *steal*) vs. ([ʃp] and ([ʃt]) (G *sprechen*, *stehlen*).

However, in other respects Dutch and German are similar: Prevocalic initial alveolar fricatives have become voiced (D *zeven*, G *sieben* ([z]) vs. E *seven*), and all word-final plosives are voiceless (D *brood* ([t]), G *Brot* vs. E *bread*).

Features peculiar to Dutch are the vocalization of preconsonantal [l] after a short vowel (D *koud* vs. E *cold*, G *kalt*), the change of initial Germanic [g] to [x] (D *geven* ([x]) vs. E *give*, G *geben*), the change of initial [sk] to [sx] rather than [ʃ] (D *schieten* [sx] vs. E *shoot*, G *schießen* ([ʃ]), the change of [-ft] to [-xt] (D *zacht* vs. E *soft*, G *sanft*), and the simplification of [-ks] to [-s] (D *vos* vs. E *fox*, G *Fuchs* [ks]).

History

Old Dutch or Old Low Franconian, conventionally dated from 700 C.E. to 1150 C.E., is attested only in a few, mainly fragmentary texts, but it already shows most of the previously mentioned typically Dutch features.

Thousands of texts date from the Middle Dutch period, c. 1150 C.E. to 1500 C.E., mainly produced in the provinces of Flanders, Brabant, and Holland and consisting of literary works and official documents. There was as yet no standardized language, and these

texts are written in a variety of dialects belonging to five main groups. A case system of nominative, accusative, genitive, and dative is discernible in nouns, adjectives, articles, and pronouns, but in the course of the period this system was eroded and greater use came to be made of prepositions and of fixed word order. Nouns showed the three genders of masculine, feminine, and neuter.

Modern Dutch is reckoned to date from c. 1500 C.E. The 17th and 18th centuries were marked by continued literary production; by the increased use of Dutch in political and scientific domains; by the beginnings of the development of a standard language; by the writing of grammars and dictionaries; and by the magisterial Bible translation of 1637, the *Statenbijbel* (States Bible). The 19th century saw the production of prescriptive grammars, contributing to an acceptance of a standard language, and a series of *Nederlandsche Taal- en Letterkundig Congressen* (Dutch Language and Literary Congresses), beginning in 1849, at which writers and scholars from all parts of the language area could meet to discuss the shape of Dutch without regard to political boundaries. This particular activity finally culminated late in the 20th century with the publication of the authoritative *Algemene Nederlandse Spraakkunst* (General Dutch Grammar) of 1984 (second edition, 1997) and the vast *Woordenboek der Nederlandsche Taal* (Dictionary of the Dutch Language), completed (in pre-1947 orthography) in 1998.

Phonetics and Phonology

The consonantal phonemes of Standard Dutch in careful educated speech of the western Netherlands are /p b t d c k g ʔ f v s z ʃ ʒ x h m n ŋ r l v j/. In this accent /p, t, k/ are unaspirated and voiceless; /b, d/ are fully voiced; /t, d, s, z, n, l/ are laminal alveolar consonants; /s, z/ have low-pitched friction; /x/ is [χ]; /h/ is [ɦ]; prevocally /r/ is realized as an alveolar tap [ɾ] and postvocally as an alveolar fricative [ɹ] (or approximant [ɹ]) or as a palatal fricative [j] (or approximant /j/); and /v/ is realized prevocally as a labiodental approximant [v] and postvocally as a bilabial approximant [β].

Dutch is characterized by assimilation and elision in connected speech. A voiceless plosive becomes voiced under the influence of an adjacent voiced one (*blijkbaar* /'bleɪgba:r/ 'evidently'); a voiced fricative becomes voiceless under the influence of an adjacent voiceless one (*laf zijn* /laf sɛɪn/ 'to be a coward'); and, in sequences of plosive preceded or followed by a fricative, the plosive determines the voicing (*afbellend* /'avbɛlə/ 'to ring off,' *opvouwen* /'ɔpfoʊvə/ 'to fold up'). Both regressive and progressive assimilation

are therefore common. A sequence of two identical consonants is simplified to a single one (*doos zeep* /do:se:p/ 'box of soap').

Alveolar consonants may coalesce with a following /j/, resulting in postalveolar or (pre)palatal sounds (*katje* /'kacə/ 'kitten,' *oranje* /o'rɑŋə/ 'orange,' *meisje* /'meɪʃə/ 'girl'). Within words, a glottal stop /ʔ/ is inserted after /a:, ə/ before syllable-initial vowels (*beamen* /bə'ʔa:mə/ 'to confirm').

Because /c, g, ʃ, ʒ, ɲ/ appear only as a result of assimilation and/or in words clearly felt to be foreign borrowings, such as *goal*, *chef*, and *jury*, some scholars prefer to deny them full phonemic status and treat them as marginal phonemes or as allophones of other phonemes. Similarly, the complete predictability of the appearance of /ʔ/ raises the question of whether it should be included in the list of phonemes.

The vocalic phonemes are /i ɪ y ɛ e: ɔ: ə a: ɑ o: ɔ u eɪ œy ʌu/. /e: ɔ: o:/ tend to be realized as closing diphthongs. Additional long vowels /i: y: ɛ: œ: ɔ: u: ɛ̃: ɑ̃: ɔ̃:/ appear only in recently borrowed foreign words, mainly from French, and are regarded as marginal phonemes in Dutch.

Morphology

The inflectional and derivational processes of Modern Dutch remain typically Germanic and are in principle little different from those of Modern English. Modern Dutch has lost most of the inflectional endings of the earlier stages of the language.

Nouns and articles show no case distinctions (apart from occasional relics of a genitive) and pronouns have only two, as in English. The majority of nouns form their plural by adding either *-s* or *-en* (*bakker/bakkers* 'baker/bakers', *boek/boeken* 'book/books'). There remain only vestigial traces of the distinction between the indefinite and definite declension of adjectives (*een groot boek* 'a big book,' *het grote boek* 'the big book').

Apart from *hebben* 'have' and *zijn* 'be,' verbs show only three forms in the present and two in the past tense. In common with most other Germanic languages, there is a distinction between weak verbs (D *vissen*, *ik viste*, *ik heb gevist*; E *fish*, *I fished*, *I have fished*) and strong ones (D *kiezen*, *ik koos*, *ik heb gekozen*; E *choose*, *I chose*, *I have chosen*).

As in English and German, there are derivational prefixes and suffixes that indicate the relationship between countless sets of root-related lexemes, for example *meester* 'master,' *overmeesteren* 'to overpower'; *hooren* 'to hear,' *gehoor* 'audience'; *blind* 'blind,' *blindheid* 'blindness'; and *ontploffen* 'to explode,' *ontploffbaar* 'explosive.' Compound nouns are common; they may look lengthy but the principle is

no different from that of English (e.g., *postzegelverzameling* ‘postage-stamp collection’).

Syntax

Genders of inanimate objects have been reduced to two, common and neuter, revealed syntactically in the choice of definite article, *de* and *het*, respectively, and of pronominal reference, *hij/hem* and *het*, respectively; thus, *de stoel* (common) ‘the chair’ = *hij/hem* ‘he/him,’ *de tafel* (common) ‘the table’ = *hij/hem* ‘he/him,’ *het boek* (neuter) ‘the book’ = *het* ‘it’ (but see the section on Regional and Social Variation). The pronouns *hij/hem* and *zij/haar* are used to refer to males and females, respectively; *de man* ‘the man’ = *hij/hem*, *de vrouw* ‘the woman’ = *zij/haar* ‘she/her,’ *het meisje* ‘the girl’ = *zij/haar*. Moreover, in the written language *zij/haar* and the possessive *haar* are used to refer to collective nouns denoting people (e.g., *de jeugd* ‘youth’ as in the sample sentence).

In declarative main clauses, the finite verb appears in second position and may be preceded by the subject or some other element, thus *ik zag hem gisteren* ‘I saw him yesterday.’ If the initial position is not occupied by the subject, the subject follows the verb, thus

| | | | |
|-----------------------|-----|----|-----|
| gisteren | zag | ik | hem |
| yesterday | saw | I | him |
| ‘yesterday I saw him’ | | | |

A past participle or infinitive appears at the end of the clause, thus

| | | | | |
|-----------------------------|------|----|-----|--------|
| gisteren | heb | ik | hem | gezien |
| yesterday | have | I | him | seen |
| ‘yesterday I have seen him’ | | | | |

In subordinate clauses, the verb is in final position:

| | | | | | | | | |
|---------------------------------|----|-----|------|------|----|-----|-----|--------|
| als | ik | hem | zie, | zal | ik | het | hem | zeggen |
| if | I | him | see, | will | I | it | him | tell |
| ‘if I see him, I will tell him’ | | | | | | | | |

Here there is obligatory inversion of subject and verb in the main clause because the subject does not begin the sentence.

In *wh*-questions, the *wh*-element occupies first position and the finite verb the second: *waar is het station?* ‘where is the station?’ In yes/no questions, the finite verb occupies first position and the subject the second:

| | | | | | |
|----------------------------------|-----|-------|-----|-------|----------|
| heeft | de | vrouw | het | huis | gekocht? |
| has | the | lady | the | house | bought |
| ‘has the lady bought the house?’ | | | | | |

Many verbs have a stressed prefix in the infinitive (e.g., *ondergaan* ‘to set’). In the present and past

tenses, this prefix **follows** the verb and appears at the end of main clauses:

| | | | | | | |
|----------------------------|-----|--------|----|-----|--------|-------|
| de | zon | gaat | in | het | westen | onder |
| the | sun | V.PRES | in | the | west | PRT |
| ‘the sun sets in the west’ | | | | | | |

But in subordinate clauses it appears as a prefix: *als de zon ondergaat, wordt het donker* ‘when the sun sets, it gets dark.’

Vocabulary

The vocabulary of Dutch is basically Germanic and has preserved hundreds of words that are not found in English (e.g., *gesprek* ‘conversation’). Nevertheless, it has many Latin and Romance, mainly French, borrowings (e.g., *straat* ‘street’ and *horloge* /hɔːrˈloːʒə/ ‘watch’). Since World War II, a large number of English borrowings have appeared (e.g., *management*, *website*), as well as loan translations (e.g., *diepvries* ‘deep freeze’).

Orthography

The present-day spelling system was basically established in the 19th century and a reformed version was made official in 1947. It is to a great extent phonemic, although not entirely so; thus, in *kat* ‘cat,’ *brood* ‘bread,’ and *wordt* ‘becomes’ the <t>, <d>, and <dt> all represent /t/. Furthermore, long and short vowels are distinguished in closed syllables by the use of two vowel letters versus one (*brood* /broːt/ ‘bread’ but *klok* /klɔk/ ‘clock’) and in open syllables by the use of one consonant letter versus two (*broden* /ˈbroːdə/ ‘loaves’ but *klokken* /ˈklɔkə/ ‘clocks’). The digraph <oe> is used for /u/ (*boek* /buk/) and <ui> for /œy/ (*huis* /hœys/), <w> represents /v/ (*wind* /vɪnt/ ‘wind’), and both <ch> and <g> represent /x/. Final <n> in place names and plurals of nouns and verbs is not pronounced in normal speech. The assimilation and elision typical of connected speech are not reflected in modern Dutch spelling.

Sample Sentence

This example is from Huizinga (1939, repr. 1950, 313).

| | | | | | |
|-----|--------|-------------|----------|--------|------------|
| de | meeste | bloeiende | culturen | hebben | wel |
| /də | mɛːstə | ˈbluːjəndə | kœlˈtyrə | ˈhebə | ʋɛl |
| the | most | flourishing | cultures | have | admittedly |
| de | jeugd | liefgehad | | en | |
| də | ˈjœxt | ˈlifxəɦat | | ən | |
| the | youth | loved | | and | |

vereerd, maar haar niet gecajoleerd of gefêteerd,
 və're:rt mar ha:r nit xəkəʒə'le:rt ɔf xəfə'te:rt
 honored but her not flattered or fêted,

en steeds van haar geëist
 ɔn stɛ:ts vən ha:r xə'ʔist
 and always from her demanded

gehoorzaamheid en eerbied voor de ouderen
 xə'ho:rzɑ:mhɛit ɔn 'e:rbit vɔ:r də 'ʌudərə/
 obedience and respect for the elderly

'while most flourishing cultures have loved and
 honored the young, they have not indulged or
 spoilt them, and always required from them
 obedience and respect for their elders'

Regional and Social Variation

The Dutch linguistic area includes not only the standard language but numerous regional variants; as many as 28 dialects, falling into six main groups, have been recognized. These dialects are dying out, however, and are being replaced by regionally colored varieties of Standard Dutch.

There is much variation in the educated pronunciation of Standard Dutch; a particularly salient set of differences distinguishes pronunciations to the north of the Rivers Rhine, Meuse, and Waal from those to the south. In the south, an additional phoneme, a voiced velar fricative /ɣ/ (spelled <g>) appears, and the corresponding voiceless fricative /x/ (<ch>) is velar rather than uvular. In fact, in the north, /v/ may be replaced by /f/ and even /z/ by /s/. In the south, particularly Belgium, /e: ɔ: o: / are monophthongs rather than closing diphthongs, and /v/ is [β] or (with palatalization) /ɥ/. /r/ appears in various realizations, usually alveolar in Belgium, Amsterdam, and the northeast Netherlands, but appears as uvular trills, fricatives, or approximants elsewhere.

In 1973 the official name of the language in Belgium became 'Dutch' rather than 'Flemish', the standard form being held to be that of the Netherlands, and in 1982 the Nederlandse Taalunie (Dutch Language Union), a joint Belgian-Netherlands venture, was set up with the aim of advancing Dutch language and literature throughout the Dutch-speaking area. Nevertheless, there are a few minor differences at the grammatical and lexical level between Standard Dutch in Belgium and in the Netherlands.

In Belgium, there are still three genders, revealed in pronominal reference as well as in the choice of article; thus *de stoel* (masculine) 'the chair' = *hij/hem* 'he/him,' *de tafel* (feminine) 'the table' = *zij/haar* 'she/

her,' *het boek* (neuter) 'the book' = *het* 'it.' A further striking difference is that the informal pronoun of address in the Netherlands is *jij/je*, whereas in Belgium it is *gij/ge*, a form that in the Netherlands is now reserved for the Deity.

Lexical differences are found not only in administrative terms referring to the different political structures of Belgium and the Netherlands, but also in a long heterogeneous list of colloquial words; thus, Belgian *blokken* but Netherlands *studeren* 'to study,' Belgian *solden* but Netherlands *uitverkopen* 'bargains, the sales,' Belgian *gans proper* but Netherlands *helemaal schoon* 'very nice,' and so on.

Dutch is no different from other languages in possessing various sociolects (varieties of languages spoken by particular social classes or ethnic, age, employment, or religious groups) and in undergoing constant change. One innovation, noticed in the last quarter of the 20th century and termed Polder Dutch, is the change in the realization of the diphthongal phonemes /ɛi œy ʌu/ from [ɛi œy ʌu] to [ai ɔu ʌu]. This began among educated, upper-class women and rapidly spread to other groups, including men and children, and can now be found throughout the Netherlands.

Influence on Other Languages

Dutch was the official language of the colonial empire of the Netherlands and various local and creolized forms of the language sprang up in present-day Indonesia, the Caribbean, and South America. Most of these are now extinct, although Dutch had a considerable influence on the still-extant Spanish- and Portuguese-based Papiamentu and English-based Sranan. The Dutch spoken by colonists sent to the Cape by the Dutch East India Company in the 17th century evolved into Afrikaans, now one of the official languages of South Africa.

Within the Netherlands, Frisian is strongly influenced by Dutch, particularly in its vocabulary (for example *sleutel* 'key' instead of *kaai*); in Belgium, the conversational French of bilingual Dutch and French speakers may show Dutch influence (e.g., *s'il vous plaît!* instead of *voilà!* when giving someone something, on the model of *alstublieft!* 'please!').

In the late Middle Ages, trade between the British Isles and the Low Countries brought Dutch words into Scots (e.g., *pinkie* 'little finger') and English, especially nautical terms (e.g., *deck*, *smuggler*, and *yacht*). The 17th century saw the introduction of artistic terms (e.g., *easel*, *landscape*, and *sketch*). More modern borrowings include *boss*, *coleslaw*, and *cookie*.

Bibliography

- Belemans R & van Thienen J (1998). *Ich kal ooch Limburgs*. Hasselt: Provinciebestuur van Limburg.
- Booij G (1995). *The phonology of Dutch*. Oxford: Oxford University Press.
- Booij G (2002). *The morphology of Dutch*. Oxford: Oxford University Press.
- Collins B & Mees I (1982). 'A phonetic description of the consonant system of Standard Dutch.' *Journal of the International Phonetic Association* 12, 2–12.
- de Vries J W, Willemyns R & Burger P (1995). *Het verhaal van een taal*. Amsterdam: Bert Bakker/Prometheus.
- Donaldson B C (1981). *Dutch reference grammar*. Groningen: Walters Noordhoff.
- Donaldson B C (1983). *A linguistic history of Holland and Belgium*. Dordrecht: Kluwer.
- Donaldson B C (1997). *Dutch: comprehensive grammar*. London: Routledge.
- Gussenhoven C (1999). 'Dutch.' In *Handbook of the International Phonetic Association*. Cambridge, UK: Cambridge University Press. 74–77.

- Haeseryn W, Romijn K, Geerts G, de Rooij J & van den Toorn M C (1997). *Algemene Nederlandse Spraakkunst* (2nd edn.). Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
- Huizinga J (1935). *In de schaduwen van morgen*. Repr. in Huizinga J (1950). *Verzamelde werken VII*. Haarlem: H.D. Tjeenk Willink en Zoon N.V. 313–429.
- Mees I & Collins B (1983). 'A phonetic description of the vowel system of Standard Dutch.' *Journal of the International Phonetic Association* 13, 64–75.
- Stroop J (1998). *Poldernederlands; waardoor het ABN verdwijnt*. Amsterdam: Bert Bakker/Prometheus.

Relevant Website

- <http://www.ned.univie.ac.at/publicaties/taalgeschiedenis> – History of the Dutch Language. Department of Dutch Studies, University of Vienna.

This page intentionally left blank

E

Eblaite

A Archi, University of Rome, Rome, Italy

© 2006 Elsevier Ltd. All rights reserved.

Eblaite, the Semitic language spoken by the people responsible for the urbanization of northern Syria and northern Mesopotamia in the early part of the third millennium B.C., is named after the city of Ebla (c. 60 km south of Aleppo) because of its archives, which cover, however, only the second half of the 24th century B.C. Eblaite is a branch of Akkadian, which dates to the 26th century B.C., whereas the texts of the Dynasty of Akkad appeared a few decades after the documents of Ebla. Eblaite was not a written lingua franca used to communicate in a large area, because the material present in the various types of sources reflect a single language. Forms with the pattern *parrus*, *parrusum* are common to Eblaite and the Assyrian dialect, whereas Old Akkadian follows the *purruš*, *purrušum* pattern.

Sources

The archives of Ebla included originally about 3000 clay tablets in cuneiform writing. Most of the texts are administrative in character and relate to palace activity. They make large use of Sumerian logograms for substantives and verbs (according to the archaic use of cuneiform); therefore, the Semitic elements include a few names of objects, most of the prepositions, and personal names. Personal names are about 3000 and geographic names about 900. The chancery documents (approximately 60), which include letters, royal decrees, some political agreements, and diplomatic reports, are richer in Semitic elements. One of the several Sumerian lexical lists was provided for most of its 1200 words with an Eblaite translation; another list of 330 words was added to it. About 50 administrative documents of the same period come from Mari (Middle Euphrates); other 200 were found in Nabata (Tell Beydar) in northern Habur.

Grammar Features

Phonological and morphological interpretation is often hampered by the inaccuracy offered by syllabic orthography, notwithstanding the rules fixed in the 'syllabary.'

Personal Pronouns

1st sg. nom. *'anna* (Akkadian *anāku*; Ugaritic *'nk*); 2nd sg. m. *'anta*, gen.-acc. *kuwāti* (O. Ass. *ku(w)āti*), dat. *kuwāši*; 3rd sg. m. nom. *šū* / *šuwā*, gen.-acc. *šuwāti*, dat. *šuwāši*. *Personal pronoun suffixes*: 1st sg. com. *-ī* / *-ya*; 2nd sg. gen.-acc. m. *-ka*, f. *-ki*, 2nd sg. m. dat. *-kum*; 3rd sg. m. gen.-acc. *-šu*; 3rd sg. m. dat. *-šum*; 1st pl. gen. *-nā* / *-nū*; 3rd pl. m. gen.-acc. *-šunū*. *Relative pronouns*: sg. m. nom. *šu*, gen. *ši*, acc. *ša*, fem. nom. *šātu*, gen. *šāti*; dual gen.-acc. *šā*; pl. gen.-acc. *šūti*, fem. *šāti*. *Interrogative pronoun*: animate nom. *mannu*, acc. *manna*; inanimate nom. *mīnu*, acc. *mīna*.

Nouns

Only Eblaite and Akkadian, among the Semitic languages, present the entire nominal inflexion of case, including the use of the dative and the locative: sg. m. nom. *-um*, gen. *-im*, acc. *-am*, dat. *-iš*, loc. *-ūm*, fem. nom. *-atum*, gen. *-atim*; dual nom. *-ān*, gen.-acc. *-ayn*; pl. nom. m. *-ū*, gen.-acc. *-ī*.

Prepositions

in 'in,' *ana* 'to,' *ašta/i/u* 'from, with, by' constitute three important isoglosses with Akkadian. *sin* 'for, toward' might be found also in Sabaeen. *al* 'on,' *min* 'in,' *minu* 'from,' *ade* 'instead of,' *qidimay* 'before,' *balu/i* 'without' are Common Semitic. The conjunction *šumma* 'if' is an isogloss with Akkadian and Arabic. *ap* 'further, rather' occur also in Ugaritic and Hebrew.

Verbs

There are three tenses: the imperfect with prefixes and some suffixes; the perfect with suffixes but

no prefixes; the imperative. The roots can be modified in common with other Semitic languages. Eblaita has G, Gt, D, Dt, Š, Št (but no N) conjugations. The prefix vowel of the 3rd m. sg. is *i-*, as in Akkadian; some verbal forms in personal names present also *ya-*, as in West Semitic. The 3rd fm. sg. has not only *ta-* but also *ti-*.

Bibliography

- Archi A (2002). 'Prepositions at Ebla.' In Gordon C H & Rendsburg G A (eds.) *Eblaitica* 4. Winona Lake, IN: Eisenbrauns. 1–21.
- Conti G (1990). 'Il sillabario della quarta fonte della lista lessicale bilingue eblaita.' In Fronzaroli P (ed.) *Miscellanea eblaitica*, 3. Firenze: Università di Firenze. 1–220.
- Fronzaroli P (1982). 'Per una valutazione della morfologia eblaita.' *Studi Eblaiti* 5, 93–120.
- Fronzaroli P (1987). 'Le pronom déterminatif-relatif à Ebla.' *Mari Annales de Recherches Interdisciplinaires* 5, 267–274.
- Fronzaroli P (1990). 'Forms of the dual in the texts of Ebla.' *MAARAV* 5–6, 111–125.
- Krebernik M (1982). 'Zu Syllabar und Orthographie der lexikalischen Texte aus Ebla: Teil 1.' *Zeitschrift für Assyriologie* 72, 178–236.
- Krebernik M (1983). 'Zu Syllabar und Orthographie der lexikalischen Texte aus Ebla: Teil 2.' *Zeitschrift für Assyriologie* 73, 1–47.
- Krebernik M (1988). *Die Personennamen der Ebla-Texte*. Berlin: Dietrich Reimer Verlag.
- Krebernik M (1996). 'The linguistic classification of Eblaita: methods, problems, and results.' In Cooper J S & Schwartz G M (eds.) *The study of the Ancient Near East in 21st century*. Winona Lake, IN: Eisenbrauns. 233–249.
- Pettinato G (1982). *Testi lessicali bilingui della biblioteca L. 2769. (Materiali Epigrafici di Ebla – 4)*. Napoli: Istituto Universitario Orientale.

Efik

B Connell, York University, Toronto, ON, Canada

© 2006 Elsevier Ltd. All rights reserved.

Background

Efik is one of the better-known African languages and was at one time one of the best-described African languages. It is spoken today by perhaps 750 000 people as a first language in the southeastern corner of Nigeria, in and around the city of Calabar, its cultural center. Due to its location near the Atlantic coast, Calabar and the Efik were encountered early by European explorers, traders, and missionaries. Calabar was a major slave port during the era of the trans-Atlantic slave trade. As a result of its strategic location, Efik became the dominant language of the region, and for a considerable period was used inland along the Cross River as the local trade language. As a result of missionary activity in the mid- to late-1800s, Efik became one of the first languages of sub-Saharan Africa to be reduced to writing. A sketch grammar was published in 1857 (Goldie, 1857), followed by the first Efik-English dictionary in 1862 (Goldie, 1862). Insights from Efik data were important in the development of our current understanding of tone and its place in phonology (Ward, 1933; Welmers, 1959; Winston, 1960), leading to the notion of downstep and indirectly to the advent of autosegmental

theory. In recent years Efik has been eclipsed in prestige locally, and it has given way to English and Nigerian Pidgin as lingua francas in Southeastern Nigeria. Despite its early development, only a small literature has accrued in the language; it is used to some extent on radio and television, but there is no Efik language newspaper. Goldie (1857) remains the only attempt at a grammar of the language, and except for the work of T. L. Cook (Cook, 1969, 1985, 1986, 2002), it has largely escaped the notice of contemporary linguists.

Classification

Efik is now recognized as part of Lower Cross, a subgroup of Cross River, which is in turn a branch of Benue-Congo and part of the Niger-Congo phylum (Connell, 1994). Linguists working on the classification of African languages have frequently noted similarities between Efik and Bantu, (e.g., Guthrie, 1967–1971; Greenberg, 1963; see also Winston, 1970), though without advancing the claim that Efik or its sister Lower Cross languages are themselves Bantu or Bantoid. Other members of the Lower Cross grouping include Ibibio and Anaang, which, despite interesting structural differences among the three, exhibit a fair degree of mutual intelligibility, Ekit, Oro, Obolo, Usaghade, and a

number of other smaller languages. The variety of Efik spoken at Calabar is considered the standard form of the language; there are minor dialect variations, spoken at neighboring Creek Town and to the northwest of Calabar in the Odukpani area. The brief descriptive notes that follow are confined to remarks on certain interesting aspects of the Efik tone system.

Efik Tone

Efik is a classic example of a two-tone (high, low) terrace-level register tone language exhibiting both 'automatic' and 'nonautomatic' downstep. Thus, after H there exist only three possibilities: H, L or downstepped H ([↓]H), and after L only H is possible. H after L and [↓]H are both lowered relative to a preceding H (i.e. 'automatic' and 'nonautomatic' downstep, respectively) and no subsequent H within the same phonological phrase can rise above this new level, hence the terracing effect. H and L can combine within the same syllable to give surface contours, both rising (LH) and falling (HL). Lexical roots in Efik fall into one of three tone classes, those bearing H, L, or LH tone patterns.

Tone functions both lexically and grammatically in Efik. At the lexical level, there is an abundance of minimal pairs that establish this function. Grammatically, tone is used to mark a wide range of functions. In the noun phrase, associative and genitive constructions, whether adjective + noun or noun + noun, are indicated by means of a modification of the tone of the second element of the construction. In the verb phrase the positive imperative form of the verb is considered to bear the inherent tone; certain tenses, aspects, and moods, and focus and negation are indicated through tone modifications. Person (2SG vs. 3SG, 2PL vs. 3PL) is marked through a modification of the tone of a prefix. Together with tense and aspect, focus in Efik may be considered a basic category of the verb system. A very brief and partial sketch only of its operation is possible here; the interested reader is referred to Cook (1985, 2002) for further details.

Focus is marked by means of an inflectional affix, normally consisting of one or more tones, attached to the verb stem. On positive verbs, a three-way distinction is made, where focus is on the verb itself (verb phrase focus, VPF), on a word or words preceding the verb (PrVF), or on a word or words following the verb (PoVF). Thus, a verb with the same time

and aspect reference will vary in its tone according to its focus condition. Verbs bearing inherent H become in the past, for 1SG, [↓]H (VPF), H (PrVF), and L (PoVF); verbs with inherent L remain L under these conditions. For negative verbs only a two-way distinction is realized, between PrVF on one hand and VPF and PoVF on the other.

This article has given only a brief insight into the complexity of one aspect of Efik, its tone system. Other aspects of Efik structure, both in the realm of phonology and in syntax, as well as elsewhere, are equally interesting. Renewed focus on this language would well repay linguists for their efforts.

Bibliography

- Connell B (1994). 'The Lower Cross languages: a prolegomena to the classification of the Cross River languages.' *Journal of West African Languages* 24(1), 3–46.
- Cook T L (1969). 'Efik (Calabar – Creek Town Dialect).' In Dunstan E (ed.) *Twelve Nigerian languages: a handbook on their sound systems for teachers of English*. London: Longmans, Green and Co. 35–46.
- Cook T L (1985). *An Integrated Phonology of Efik 1*. Ph.D. dissertation, University of Leiden.
- Cook T L (1986). 'The chameleonic vowel in the harmonizing prefixes of Efik.' In Bogers K, van der Hulst H & Mous M (eds.) *The phonological representation of suprasegmentals: studies on African languages offered to John M. Stewart on his sixtieth birthday*. Dordrecht: Foris. 209–232.
- Cook T L (2002). 'Focus in Efik.' *Journal of African Languages and Linguistics* 23(2), 113–152.
- Goldie H (1857). *Principles of Efik grammar, with specimen of the language*. Calabar: United Presbyterian Church.
- Goldie H (1862). *Dictionary of the Efik language* (2 vols). Edinburgh: United Presbyterian College.
- Greenberg J H (1963). *The languages of Africa*. Bloomington: Indiana University Press.
- Guthrie M (1967–1971). *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages*. Farnborough, England: Gregg International.
- Ward I C (1933). *The phonetic and tonal structure of Efik*. Cambridge: W. Heffer & Sons.
- Welmers W E (1959). 'Tonemics, morphotonemics, and tonal morphemes.' *General Linguistics* 4(1), 1–9.
- Winston F D D (1960). 'The "mid" tone in Efik.' *African Language Studies* 1, 185–192.
- Winston F D D (1970). 'Some Bantu-like features of Efik structure.' *African Language Studies* 11, 417–435.

Elamite

G B Gragg, University of Chicago, Chicago, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Elamite was a language spoken on the Iranian plateau from at least the end of the fourth millennium B.C.E. until at least the end of the Persian Empire in the fourth century B.C.E. It is attested in thousands of cuneiform documents, the great majority of which come from the western and southwestern periphery of that area, in the modern provinces of Khuzestān and Fārs, especially from the sites of Susa, which was an important political and economic center during the whole period, and Persepolis, the dynastic center of the Achaemenid Empire. The Elamite texts are almost exclusively either monumental (e.g., building inscriptions) or administrative (e.g., accounts, rations) in character. Elamite was first identified by European scholarship early in the 19th century because of its inclusion in many of the Achaemenid bilingual and trilingual monumental inscriptions: Old Persian, written in a specially created alphabetic script, and Akkadian (Babylonian) and Elamite, both written in Mesopotamian cuneiform. This corpus provided the basis for the decipherment of Akkadian, and hence Elamite, cuneiform in the 1840s.

The earliest evidence of texts in the region was found on about 1500 clay tablets dating from about the end of the fourth and beginning of the third millennium B.C.E. They are written in a still undeciphered script termed, misleadingly, ‘Proto-Elamite.’ The language of these texts is presumed, but not proven, to be Elamite. The same can be said about another small corpus of about 20 texts from the end of the third millennium written in an indigenous but cuneiform-like script known as ‘Linear Elamite.’ Apart from these two very early corpora, Elamite texts are written in various local adaptations of Mesopotamian cuneiform, usually with a significant reduction in the number of syllabic and logographic values and the occasional creation of a new local value. This corpus of texts is unevenly distributed over the commonly recognized periods given in Table 1. Achaemenid Elamite is obviously the best attested and most studied of these corpora, but, in part because of the massive presence of Old Persian and other Iranian names and loanwords in this period and doubts about the extent of Iranian influence on the development of the Elamite language, Middle Elamite, where it differs from Achaemenid Elamite, tends to be taken as a kind of classical norm.

Elamite cannot with any certainty be related to any known language, although a geographically plausible

relation to the Dravidian languages has been discussed almost since the beginning of modern scholarship in the language (see the synthesis of McAlpin, 1981). Partly because of the lack of a known cognate language and partly also because of the relatively limited and stereotyped nature of the contents of the corpus, the interpretation of many aspects of Elamite grammar remains uncertain and subject to a great deal of discussion in the literature.

In terms of word-order typology, Elamite has predominantly subject-first, verb-last order, with a certain amount of variability. Adjectival, genetical, and relative modifiers normally follow the head noun; adverbial relations are rendered principally by postpositions, although there are also some prepositions. It is a subject of current debate whether Elamite displays features that might be termed ‘ergative,’ but in any case it is much less so than its contemporary neighbors Hurrian, Urartian, and Sumerian. Elamite morphology is almost exclusively suffixing. A central role in morphology and syntax is played by a classifier suffix/enclitic, NCLASS, marking person, number, and various classes of animacy and inanimacy. This formative plays the role of subject marker in two of the three verb conjugations, but it is also used in one of most characteristic features in the language: the delineating of a syntactic constituent N_{HEAD} adjunct by one or more occurrences of a NCLASS agreeing with the head. The adjunct can be an adjective or a possessor, as in:

- (1) temti riša-r
lord.ANIM.SING great-NCLASS.ANIM.SING
‘great lord’
- (2) ulhi sunki-me
house.INAN.SING king-NCLASS.INAN.SING
‘house of the king’

Table 1 Periodization of Elamite

| Period B.C.E. | Designation | Texts |
|---------------|--------------------|--|
| c. 2600–1500 | Old Elamite | Less than a dozen texts of varied content |
| c. 1500–1000 | Middle Elamite | Almost 200 royal monumental inscriptions |
| c. 1000–550 | Neo-Elamite | About 30 royal inscriptions and several hundred legal and administrative texts |
| 550–330 | Achaemenid Elamite | The several hundred royal inscriptions and several thousand administrative texts from the administrative centers of the Persian Empire |

or in some more complex combination of the two:

- (3) sian Nabu-me upat-hussi-p-me kuši-h
 temple.INAN.SING Nabu-NCLASS.INAN.SING
 brick.AN.PL-baked-NCLASS.AN.PL-NCLASS.INAN.SING
 build-I
 'I built a Nabu temple of baked brick'

The adjunct can also be used as relative (note the additional NCLASS formative after the negation marker before the relativized verb):

- (4) sian in-me kuši-h-š-me-a
 temple.INAN.SING not-NCLASS.INAN.SING build-PL-
 3RD-NCLASS-COMP
 'the temple which they did not build'

Finally, this marking can occur even in sequences which, for example, in English, would not normally be treated as a syntactic constituent (on this construction see Stolper, 2004: 85):

- (5) peti-p pat-p u-p rabba-k-na
 enemy.AN.PL-NCLASS[AN,PL] under-NCLASS.AN.PL me-
 NCLASS.AN.PL bind-PERF-OPT
 'may enemies be bound under me'

The last example illustrates that the NCLASS can also occur on the head noun, in this case explicitly marking it as plural.

Bibliography

- Grillot-Susini F & Roche C (1987). *Éléments de grammaire élamite*. Paris: Editions Recherche sur les Civilisations.
- Hallock R (1969). *Persepolis fortification tablets*. Chicago: University of Chicago Press. [contains important glossary of Achaemenid Elamite.]
- Hinz W & Koch H (1987). *Elamisches Wörterbuch*. Berlin: Dietrich Reimer.
- Khačikjan M (1998). *The Elamite language*. Rome: Istituto per gli studi micenei et egeo-anatolici.
- McAlpin D (1981). *Proto-Elamo-Dravidian: The evidence and its implications*. Philadelphia: American Philosophical Society.
- Paper H (1955). *The phonology and morphology of Royal Achaemenid Elamite*. Ann Arbor: University of Michigan Press.
- Reiner E (1969). 'The Elamite language.' In Spuler B (ed.) *Handbuch der Orientalistik: Altkleinasiatische Sprache*. Leiden: Brill. 54–118.
- Stève M-J (1992). *Syllabaire élamite: histoire et paléographie*. Paris/Neuchatel: Recherches et Publications.
- Stolper M (2004). 'Elamite.' In Woodard R (ed.) *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press. 60–94.

Endangered Languages

L Grenoble, Dartmouth College, Hanover, NH, USA

© 2006 Elsevier Ltd. All rights reserved.

Language Endangerment

Although it is somewhat difficult to count languages and to measure linguistic diversity with exact precision, there are an estimated 6800 languages spoken in the world today. While there is some question as to exactly how many languages will be lost over the course of this century – ranging from a low of 25% to a high of 90% – there is widespread agreement that language loss is occurring at an unprecedented rate. Most recent studies have concluded that at least 50% of the world's languages are losing speakers and that by the end of this century, a full 90% of the world's languages will disappear entirely, replaced by more widely used (national and/or global) languages. This situation is generally referred to as language endangerment, a term used broadly for languages which are threatened with absolute loss;

a language is considered lost when it has no speakers. Language endangerment is sometimes called language attrition or language death, but 'death' is avoided out of sensitivity to the population whose language has been lost. Language attrition and moribundity – when children cease learning a language – are now taking place with exceptionally rapid speed. Hundreds of languages are currently endangered and there are few parts of the world where some form of language decline is not occurring. While language attrition is not in and of itself a new phenomenon, the rate of decline in linguistic diversity appears to be unique to this era, and is perhaps rivaled only by the kind of language loss which took place in conjunction with the agricultural revolution of approximately 10 000 years ago. One consequence is that a significant number of communities are facing the loss of a language which historically and traditionally has been foundational to their sense of identity. In some instances communities are reacting with efforts to revitalize the local language, while in others they lack the resources, time, or motivation to do so.

Linguists are particularly concerned with the loss of indigenous, or local, languages, as opposed to immigrant languages. For the latter, the language may give way in the new territory to an already established (national or dominant) language, but a robust speaker community continues/thrives in the homeland of the immigrants. (This is the situation of most immigrant languages in the United States, for example; for the most part, second-generation immigrants speak and use English in their daily lives, but their ancestral language is maintained in their original homeland. English, in contrast, is an immigrant language to North America; for a range of historical, socio-economic, and political reasons, it has largely ousted Native American local languages.) It is the loss of such local languages which is of concern to linguists, as their loss means an absolute kind of disappearance of the language. Thus, by and large, the term 'language endangerment' refers to the attrition and potential loss of local languages. A language is considered endangered when it is used by fewer speakers and when it is used in fewer situations or domains.

Language endangerment typically involves language contact situations, with two (or more) languages in use, where one language (Language A) replaces another (Language B). Prototypically, Language A is being adopted by speakers of Language B and so Language A replaces Language B in the sense that decreasing numbers of speakers of Language B use it, until ultimately there are no speakers of Language B at all. This is referred to as language shift, a term which refers specifically to such changes in patterns of language use, whereby speakers abandon the language of their parents in favor of another language. In the scenario outlined here, Language A can most neutrally be referred to as a language of wider communication; it tends to be a language which holds social prestige, serves official and governmental functions, and is used in education. It is often a regional or national *lingua franca*, i.e., the language which groups speaking different languages use to communicate with one another. It is also called, less neutrally, the dominant language, the majority language, or even the killer language. 'Dominant language' is to be avoided as it implies a deliberation on the part of the speakers of that language to dominate others; in some instances this is in fact the case, as when language policies intentionally restrict use of a local language. But in other situations the influence of the language of wider communication is more indirectly and subtly attained, through prestige and social pressures. Similarly, a number of labels are used to refer to Language B, such as minority language, indigenous language, mother tongue, or

heritage language. The term 'local language' is more neutral and captures the fact that language use is tied to a particular geography, and that a speaker community generally sees the need or desire to use this language within a given region. The respective terms 'majority' and 'minority' for Languages A and B are not always accurate; speakers of Language B may be numerically greater but in a disadvantaged social or economic position which makes the use of the language of wider communication attractive. The term 'heritage language' can be confusing, as it is often used to refer to the language of one's ancestors, regardless of how many generations have passed since anyone spoke the language. It does not necessarily refer to a local or indigenous language, and can also refer to the ancestral languages of immigrants, even when they have not been spoken for generations.

Predications of language loss stem from several considerations, which center around a combination of critical factors in language vitality, including the number and generations of speakers, their geographic distribution and relative isolation, and recognition of ongoing rapid language shift. First, there is a very uneven distribution between languages and speakers, with just a handful of languages spoken by a very large percentage of the global population. According to current counts, approximately half the world's population speaks one of just 20 languages, and eight languages (Mandarin [Mandarin Chinese], Spanish, English, Bengali, Hindi, Portuguese, Russian, and Japanese) surpass all others with over 100 million speakers. Arabic could perhaps be added to this list: the sum total of all speakers of some form of Arabic makes it the fifth largest language, with over 200 million speakers. Not all varieties of Arabic are mutually intelligible, however, and so the differences between them are more language-like than dialect-like. Yet given the total number of people who speak some variety of Arabic, it should be included in the list of major world languages. The situation is markedly different for most of the world's languages. Some 96% of all languages are spoken by just 4% of the population, and one-fourth of the total number of languages have fewer than 1000 speakers. More than half of all languages have fewer than 10 000 speakers. Although the total number of speakers is not the sole indicator of language vitality, it is certainly a very important one. A very large majority of the world's population speaks just a very few languages. More to the point is the fact that we are witnessing rapid language shift, with a small set of major or global languages gaining in terms of numbers of speakers at the expense of a vast majority of the world's languages.

Table 1 Geographic distribution of languages

| | <i>Total living languages</i> | <i>Percentage</i> |
|--------------|-------------------------------|-------------------|
| The Americas | 1013 | 15% |
| Africa | 2058 | 30% |
| Europe | 230 | 3% |
| Asia | 2197 | 32% |
| The Pacific | 1311 | 19% |
| Total | 6809 | |

Source: Grimes (2000).

Finally, it is important to note that the geographic distribution of languages is also very uneven, with the largest numbers of languages spoken in Africa and Asia, and much smaller numbers elsewhere, such as in North and South America and the Pacific. Europe has a very few languages, both in terms of raw numbers and of percentage relative to the whole. This distribution is summarized in **Table 1**.

Distribution by continent or region is only part of the story. Language density, or the number of languages per unit area, varies greatly. Papua New Guinea stands out with 820 languages; with its relatively small territory, it has the highest language density of any country in the world. In all of North America, fewer than 200 indigenous languages remain, although there were certainly hundreds of distinct languages several centuries ago. Today, only a handful of these (such as Cree, Dakota, Ojibwa, Navajo) have a hope of survival, and even their longevity is doubtful. The case of Cree is illustrative. As of the 1998 Canadian census, there was a total of 87 555 speakers of all varieties of Cree. These speakers are not monolingual, however, and they show low literacy rates in Cree (only 5–10%) but high literacy rates in a second language, usually English (75–100%). Such figures are indicative of significant language shift. The figures for Dakota are even more alarming, with fewer than 27 000 speakers in North America as a whole, and only 31 monolingual speakers (as of 1990). Basic descriptions of Dakota speech patterns note that in some communities the children and young adults do not speak Dakota or at least prefer to speak English. Again, these are all signs of ongoing and advanced language shift, leading toward language extinction.

As this may suggest, in only a very few cases is language loss due to the loss of the speaker population itself. Instead, the primary cause for language loss is language shift, when speakers cease to speak their own native tongue, the local language, in favor of the language of what is usually, politically and/or economically, the dominant culture. Such shift from the local to the language of wider communication can

occur over several generations, or even as quickly as over the course of a single generation. In many cases the oldest generation – the grandparent group – speaks the local language as their first and primary language and has limited or nonfluent knowledge of the external language of communication; in some instances they may even have no knowledge at all of the language of wider communication. In contrast, the middle generation has some knowledge but primarily uses the language of wider communication, and the youngest generation has little to no knowledge of the local language, using at best a few words or phrases, such as greetings. In cases of rapid language shift, however, these changes occur across a single generation.

Levels of Language Endangerment

Implicit to the study of language endangerment is the notion that a relatively vital language can change to a state of endangerment at some point, usually when children cease learning the language. In studying language endangerment, it is important to assess degrees of vitality versus endangerment. That said, because a large number of factors enter into each situation, it can be difficult to rank levels of endangerment. Therefore, different linguists have proposed a variety of scales, with differing numbers of stages of endangerment and different labels for each level. There is, however, widespread agreement on the ends of the scale: safe languages and extinct languages. Generally, languages are categorized with respect to endangerment on a scale of six levels: safe, at risk, disappearing, moribund, nearly extinct and extinct.

Safe: A language is considered safe when all generations use the language in all or nearly all domains. It has a large speaker base relative to others spoken in the same region and, therefore, typically functions as the language of government, education, and commerce. Many safe languages enjoy official status within nation-states, and as such tend to be held in higher prestige than other languages.

At Risk: A language is at risk when it is vital (being learned and used by people of all different age groups) without any observable pattern of a shrinking speaker base, but lacks some of the properties of a safe language: for example, it is spoken in a limited number of domains or has a smaller number of speakers than other languages in the same region.

Disappearing: A language is disappearing when there is an observable shift towards another language in the communities where it is spoken. With an overall decreasing proportion of intergenerational transfer, the speaker base shrinks because it is not being

replenished. Disappearing languages are consequently used in a more restricted set of domains, and a language of wider communication begins to replace it in a greater percentage of homes.

Moribund: A moribund language is one that is not being transmitted to children.

Nearly Extinct: A language can be considered nearly extinct when only a handful of speakers of the oldest generation remain.

Extinct: An extinct language is one with no remaining speakers.

It should be noted that sometimes an intermediate stage between 'safe' and 'at risk' is recognized, 'safe but small,' determined by languages which are otherwise safe and stable but have a relatively small speaker base. The last three levels of language endangerment given here – moribund, nearly extinct, and extinct – are characterized by a lack of intergenerational transmission; disappearing languages are characterized by a downward trend.

Many linguists would argue that any language which is not at the safe level is endangered. Furthermore, there does not appear to be a direct correlation between the level of endangerment and the anticipated rate of language attrition: some communities shift language usage more slowly, and others more quickly. That said, language endangerment is currently a pressing concern for the linguistic community precisely because rapid attrition is occurring at a global level. In addition, the kinds of documentation and revitalization efforts needed are directly related to the level of endangerment. The closer a language is to extinction, the greater the urgency to act before fluent speakers are gone. Except in cases of sudden attrition (e.g., when a language is abruptly lost through natural catastrophe or warfare; see the next section), endangered language situations tend to be characterized by speakers of differing proficiency levels. Languages ranked at any level below safe tend to have communities which include individuals who are semispeakers, i.e., not fully fluent speakers, lacking native proficiency; the ratio of semispeakers to fluent speakers varies among communities and with endangerment levels. Such semispeakers show varying degrees of fluency, ranging from strong or nearly fluent speakers, through reasonably fluent semispeakers and weak semispeakers who are less fluent, to those with even more limited speaking competence. In assessing language vitality, it is thus important to take into consideration the proficiency and knowledge of the speakers of the language. Even in the case of extinct languages, there may be cause to move quickly, as there may still be 'rememberers' of the language who have some recollection of its use or

may have some experience with it. Sometimes communities opt to resurrect (or 'resuscitate') extinct languages; rememberers can play a critical role in these efforts. Here too there is a range of proficiency: in some cases such rememberers have memorized entire texts without understanding their meaning, while the knowledge of others is restricted to only a few words or phrases. They can play an important role in language revitalization and documentation efforts, but they are static and do not represent living language.

What Is Lost?

There are a number of reasons to be concerned about language attrition. Language is a key part of each person's identity and is an essential component of a group's cultural and social heritage. Local communities who have lost their language speak about it as a deeply personal loss which is accompanied by a loss of a sense of self. Speakers whose languages are not endangered are also aware of the importance of language as a marker of identity and pay great attention to differences in dialects and speech patterns. Thus perhaps one of the most compelling reasons to be concerned about language endangerment is that the speakers who lost this part of their heritage deeply regret it and grieve over it. For this reason, so many different communities around the world are currently engaged in language revitalization efforts. Some of those groups whose languages are extinct are now attempting to resurrect them from whatever records have survived, including missionary descriptions, grammars, and sometimes oral recordings.

Loss of language also means a loss of intellectual wealth. From the linguistic standpoint, as we lose languages, we lose linguistic diversity. A great many of the world's broad array of endangered languages are understudied; what little knowledge we have indicates that many are structurally very different from the languages spoken by the majority of the global population (e.g., Mandarin, English, Spanish, and so on). The languages with the most speakers, cited in **Table 1**, represent a very small portion of the world's languages typologically and genetically. Thus language loss means a decline in sources about the range of human language and its limitations. For the linguistic community, one of the challenges of language endangerment is to record and describe as many languages as possible while they are still spoken, so that we do not lose this wealth of human knowledge without record. Language loss should also be considered from the broader scientific perspective. Language encodes the range of human experience and knowledge; its disappearance entails the loss of

the skills, information, beliefs, and ideas of a people. Often this involves specific knowledge about plants and their medicinal uses. It also includes historical knowledge; preliterate societies record their histories in their oral traditions, including stories, legends, and songs which tell the history of their people, settlements, battles, and so on. Language is more than a repository for religious and spiritual beliefs; in many societies the language itself is sacred and cannot be separated from religious beliefs and practices.

Taxonomy of Endangerment Situations

Language change and loss are naturally occurring processes which have been in place as long as language itself. Every language is constantly changing over time, and eventually evolves into one or more related but different languages; for example, the modern Romance languages (Spanish, Italian, French, and so on) are related to Latin, which is no longer used as a spoken language except for religious purposes. This kind of language 'loss' is a natural and ongoing process. Linguists are more concerned, however, with the absolute loss of language, which occurs when a language disappears entirely, without descendant languages. This kind of loss comes about in several different ways. Sometimes an entire speaker community passes away due to warfare, genocide, or disease. More frequently, however, language loss is the result of language shift, when speakers cease to speak their own native (local) tongue in favor of the language of what is usually the dominant culture, dominant politically and/or economically. The time frame for such shift varies across situations; it can take place over several generations, or much more quickly. One typical pattern is that the oldest generation, the grandparents, speaks the local language as their first and primary language, the middle generation has some knowledge but uses the dominant language primarily, and the youngest generation has little or no knowledge of the heritage language, and may at most know a few words or phrases. In cases of rapid language shift, however, these changes occur across a single generation, with the parent generation speaking the local language but their children, for whatever reasons, speaking a different one.

There are a number of ways to categorize language endangerment situations. One useful taxonomy takes into account the relative rate of attrition together with its causes. This taxonomy recognizes four different categories of attrition: sudden, radical, gradual, and top to bottom.

Sudden attrition refers to language loss which occurs abruptly because of the sudden loss of its speakers due to disease, war, natural disasters, and so on. Relatively

few cases of sudden attrition have been documented, although it most probably occurred more frequently during periods of colonization, when certain indigenous groups are known to have been annihilated due to disease. In modern times, civil strife, ethnic and religious clashes, and the spread of some diseases, such as HIV, increase the chances of sudden attrition occurring in certain areas of the world.

Radical attrition is similar to sudden attrition in the sense that it comes about due to political circumstances which cause speakers to stop using their language. Such circumstances include repression and/or genocide, often occurring where groups are singled out for ethnic cleansing. (Under colonization and later, apartheid in South Africa, for example, Khoisan speakers abandoned their ethnic identity and so too their languages in order to avoid repressive measures which included genocide.) Such language shift is thus a means of self-defense or even self-preservation for speakers for whom identification with their ethnic group may lead to persecution. In these circumstances people are likely to cease speaking their heritage language abruptly.

Cases of gradual attrition are more prevalent in the world today. Gradual attrition is the relatively slow loss of a language due to language shift away from the local language to a language of wider communication. In some cases the language of wider communication is a regionally dominant language, and in others a national lingua franca. Gradual attrition often involves transitional bilingualism: as the speaker population is in the process of shift, certain groups primarily speak the local language and others the language of wider communication. Thus it is here that the clearest gradations in intergenerational transmission are to be found. Because this type of attrition is gradual, speaker communities may be unaware that it is in progress, until it is quite advanced and the local language is seriously endangered.

Bottom-to-top attrition refers to the loss of the local languages in most domains with the exception of religious and ritual practices. Languages at this level are in an advanced stage of attrition. The local language is preserved only in those contexts where its use is seen to be the most critical. This tends to be those types of context where ritualized or sacred texts are critical, and the population may view the specific language of these as sacred in and of itself. Such ritualized or ceremonial texts are often memorized. Because these tend to be very prestigious but restricted domains for a community, it can be difficult to assess the actual vitality of the language in question. In less advanced instances of bottom-to-top attrition, the language is still used spontaneously in the settings to which it has been assigned by members of the local

community. In extreme cases, the only remaining knowledge of a local language may be memorized portions of a ceremony. There are reports of communities which have retained the memorized rituals in the local language for many generations but have lost all comprehension of them.

Assessing Language Vitality

The factors involved in assessing language endangerment or vitality are complex. Language vitality is usually ranked in scalar terms on the basis of a combination of factors, in particular on numbers and generations of speakers. On one end of the scale are extinct languages which are no longer spoken at all, and on the other end are viable languages in no current threat of endangerment. In between these two extremes, a number of stages can be recognized. A healthy language with strong vitality is used with a variety of functions and in a range of settings, usually called domains. The most vital languages are used in all settings, formal and informal, official and in the home. In cases of language attrition, the local language is used in increasingly fewer domains with fewer functions as attrition progresses. Simply put, an important diagnostic in assessing vitality is the range of uses of a particular language.

Although it is often thought that the absolute number of speakers of a language is the key factor in language vitality, experts agree that in fact it is intergenerational transmission which is critical in determining it. In order for a language to be healthy, it needs to be used by future generations. When children cease learning and speaking a language, it is already endangered, even if there still exists a significant number of speakers. Intergenerational transmission does not in and of itself guarantee the safety of a language, however, as a complex set of factors are involved. These all pertain to questions of who uses the language, how, and when. In 2003 UNESCO's Ad Hoc Expert Group on Endangered Languages established a core set of nine criteria to be used in determining language endangerment:

1. Intergenerational transmission
2. Absolute number of speakers
3. Proportion of speakers within the total population
4. Trends in existing language domains
5. Response to new domains and media
6. Materials for language education and literacy
7. Governmental and institutional attitudes and policies, including official status and use
8. Community members' attitudes toward their own language
9. Amount and quality of documentation.

These nine factors are key in assessing language vitality. Variables (1)–(3) involve the distribution of speakers of the language, relative to the total number of the ethnic population as well as to generational stratification, and in absolute terms as well. Variables (4) and (5) are concerned with domains of language use; (7) and (8) with attitudes at the local and national level; and (6) and (9) are related to the kinds of material available for the language, including both pedagogical and reference materials as well as linguistic documentation. Strictly speaking, the level of linguistic documentation relates to language endangerment only insofar as ample documentation can aid language revitalization or resurrection efforts; the act of documenting a language does not directly affect its vitality.

Intergenerational Transmission

Intergenerational transmission is the single most important factor in determining a language's viability. In order for a language to remain healthy, it must be spoken by children, as they are the representatives and predecessors of future generations of speakers. For this reason, intergenerational transmission is the single most critical factor in a language's ongoing vitality. At the same time, rates of intergenerational transmission may vary between villages or speaker communities and it cannot be assumed to be uniform across a speaker population. There can be variation within a single village: it is often the case that in one family the children do not learn to speak the local language but in another they do. As this accurately suggests, overall language vitality may be uneven, higher in some communities and lower in others. A thorough analysis of language vitality requires attention to such regional variation in addition to the generational variation in transmission and use. A 10-way distinction in terms of transmission and use is proposed by Krauss (1997) to enable a more detailed means of assessing variation in speaking patterns across generations:

- a. The language is spoken by all generations, including all, or nearly all, of the children.
 - a-. The language is learned by all or most children.
 - b. The language is spoken by all adults, parental age and up, but learned by few or no children.
- b-. The language is spoken by adults in their 30 s and older but not by younger parents.
 - c. The language is spoken only by middle-aged adults and older, in their 40 s and up.
 - c-. All speakers are in their 50 s and older.
 - d. All speakers are in their 60 s and older.
 - d. All speakers are in their 70 s and older.

- d-. All speakers are in their 70s and older, and there are fewer than 10 of them.
- e. The language is extinct, with no speakers.

As this scale suggests, it is important to make distinctions across age-groups within a single generation as well as across generations. Some might argue that a language is already in danger at stage (a-), when some of the children are not learning it. At stage (b), there is a greater level of danger, and so on; if the language is to survive at this stage, efforts need to be made at revitalization, or for reversing language shift. This scale may appear overly detailed; it is clear that a language is already on the way to extinction when it has reached stage (b). Yet at times it is needed. First, it can be quite useful in assessing the relative vitality not only of different languages, but at times more importantly, of various speaker communities. Inuktitut, for example, can be rated at level (a) in Greenland, where there are fluent speakers of all generations. (There are other factors which enter into its vitality in Greenland, such as official language status and use in education.) In some other Inuktitut-speaking communities, however, children are not learning the language and it is on the path to extinction. This is the case in specific communities in Canada and Alaska, although the children are acquiring it in other communities. This example further underscores the fact that evaluating the overall status of a language can be difficult, as it may vary from community to community. Second, if community members decide to revitalize their language, it is important to have an accurate understanding of the ages and numbers of fluent speakers who can assist in the revitalization effort.

Absolute Numbers of Speakers

Absolute population size alone is not a definitive indicator of language vitality. Each individual community is embedded in a set of circumstances that affect language use, so that even a small isolated rural community which has little contact with speakers of other languages and in which all members, of all generations, learn and use the local language, cannot reasonably be called endangered. That said, as a general rule, the more speakers, the more likely the community will be able to resist language shift. Put differently, small communities are at greater risk, because they can more easily disappear due to any one of a number of natural or man-made catastrophes. Furthermore, a small community can more easily be assimilated to a large community, and is likely to have fewer resources to resist external pressures. Yet small size alone does not condemn a language to extinction, because the nexus of relevant

factors may actually favor language use. A case in point is Icelandic. It is spoken as the first language by a relatively small group of people (approximately 300 000), but it is the national language of the country of Iceland, has a long-standing literary history, and is a language of education. Icelanders have a strong sense of pride in their cultural and linguistic heritage and teach Icelandic to their children as their first language. It is hard to characterize it as being in any way endangered. By the same token, a relatively large speaker community does not guarantee language vitality. Navajo, an Athapaskan language spoken in North America, provides an example. Although there are currently approximately 178 000 speakers (2000 census), there is ample evidence of advanced language shift. In 1968, a full 90% of first-grade children spoke Navajo as their first and primary language; by 1990 this figure had dropped to 30%. Despite the relatively large speaker base, it is doubtful that future generations will speak Navajo unless measures are undertaken to assure its continuance.

Proportion of Speakers within the Total Population

The ratio of speakers of the local language with respect to the total population of the local community is an important diagnostic in evaluating language vitality. For safe languages, all of the population speaks the language. In contrast, for extinct languages, none of the population does. In between these two extremes, languages can range from unsafe, with early language shift, where nearly all of the population still uses them, to severely endangered, where only a small percentage do. The larger the percentage speaking and using the language in an active way, on a daily basis, the more likely the language is to maintain its vitality.

In addition to the ratio of speakers of the local language to the number of people who would claim that local language as ancestral, it is useful to compare how the local language speakers are embedded in a larger social and cultural context. Often local communities are in a minority position with regard to a national culture, represented by speakers of a language of wider communication. The narrower the gap, the stronger the position of the local language.

Trends in Existing Language Domains

A vital language continues to be used in existing domains, while in contrast an endangered language is used in fewer domains. The differences in usage can be placed on a continuum, with safe languages used in all domains for all purposes. Next are situations

of what is called diglossia, or the use of different varieties in different contexts. Here a language of wider communication, usually a regional or national language, is the one used in official domains, such as government, education, and other public offices and institutions. The local language, in contrast, is also used in public domains, including not only traditional (local) places of worship or other religious institutions, but elsewhere as well. Typically, the local language is used in the home and informal domains, and the language of wider communication in official domains, and both can be used in public domains. Older members of the community may use only the local language. Next on the continuum is what the UNESCO Ad Hoc Group of Experts terms dwindling domains, when use of the language of wider communication spreads at the expense of the local language. The critical change here is that the local language is used less frequently in the home and is not transmitted to children. This state is further characterized by bilingualism in both the parent and grandparent generations; the children tend to be semispeakers but may be bilingual if the local language is spoken in the home. As this description suggests, at this point there is advanced language shift; the language is endangered and could be ranked as disappearing or even moribund. There are two final stages in this continuum which precede extinction: limited or formal domains is one, and highly limited domains the other. The former is characterized by use of the local language at festivals and ceremonies, in particular when the older generation is present (and therefore using the language). Often the use of the local language is itself tied to the rituals of these occasions and to an extent may be formulaic in usage. UNESCO also includes use of the language in the home in this category when such use is limited to the grandparent generation. The next stage, highly limited domains, represents even greater restriction in use of the local language. It is used only by a very small number of people on very particular occasions, and its use is often ceremonial.

As is clear from this description, the range of domains in which a language is used can be correlated to the generational distribution of speakers and their levels of proficiency. Use in all domains requires fluent speakers of all ages. Loss of intergenerational transmission, by its very definition, is indicative of a restriction in domains, as it signals that the language is not spoken in the home setting with children.

Response to New Domains and Media

Vital, safe languages are not only used in existing domains, but a measure of their strength is the extent

to which their use is extended to new domains. These are created as society and conditions change, and an important measure of a language's vitality is the extent to which it evolves with the people who speak it. The general pattern, worldwide, is for the language of wider communication to be used in emerging domains, including formal education and media of all kinds. The question of language use in the media is critical. The media helps spread language use and fosters its growth and/or maintenance. Moreover, use of a language in media is an important indication of that language's prestige and the kind of support it receives from the larger (dominant) culture, the allocation of resources, and so on. Finally, the media represents prestige and affluence, and the language used in the media is associated with both of these.

Education is a key domain for language use. By its very nature, education promotes the language of instruction and fosters its use. Many local languages are not used in schools; in places where they are, they are more likely to be taught as a secondary subject and not used as languages of instruction. For a language to be truly vital, not only must it be taught in the schools, but it also must be used to teach other subjects.

Materials for Language Education and Literacy

Most linguists and local community members agree that education and literacy in the local language are necessary to maintain vitality, or to revitalize a language threatened with endangerment. Some local communities reject this notion, wanting to preserve their oral traditions and to rely solely on them. There is, however, a cost to this decision, as it limits the domains in which the language can be used. Regardless, most regard literacy as essential for local languages. Yet more than half of all languages have no written form, and so a writing system needs to be developed for them in order to use them in education and literacy programs. Basic pedagogical and reference materials are needed, including textbooks, dictionaries and usable descriptive grammars. Such materials are readily available for languages of wider communication, but not for the majority of local languages. In addition, reading material is needed for literacy as well.

The existence and use of such materials is another diagnostic for assessing language vitality. UNESCO uses a scale of six levels in this assessment; each of these levels correlates with ever-decreasing vitality. At one end, safe and stable languages have an established orthography with a written tradition that includes a full range of written materials; the language is used in official domains such as government and education.

At the next level, the materials exist and are used by children in the school, at least in terms of developing local language literacy, but the written language is not used in the government or administration. At the third level, children are exposed to written materials in the schools; they may play a role in education but print media, such as newspapers, journals, magazines, do not use the written form of the language. At the next level, although written materials exist, they are not used in education. Only some community members use them, while for others, their existence may have symbolic value. At the fifth level, the community has knowledge of a writing system and some written materials exist. Finally, at the other end of the community, there is no orthography and the language has no written form.

The singular importance of literacy, as presented by the UNESCO Ad Hoc Group of Experts, is not one which would be embraced by all linguists and by all community activists. It represents a practical view of the role of writing and literacy in the modern world in which local languages compete to survive.

Governmental and Institutional Attitudes and Policies

National and regional governmental policies, laws, and attitudes all play a critical role in the fate of local languages. Policies can be viewed as supportive, fostering the use and development of local languages. They can be benign, not explicitly supportive but also not disadvantageous to local languages. Governmental policies can also be explicitly hostile toward local languages and can actively discourage their use.

There is a direct relation between national-level policies and the attitudes of speakers of the language of wider communication. Positive policies at the national level tend to reflect the overall attitudes of the population toward local languages. One aspect of this is attitudes toward bi- or multilingualism. While some nation-states (such as Canada, Nigeria, or Switzerland) are multilingual, with multiple national and/or official languages, others (such as the United States) are unequivocally monolingual at a national level with regard to language and education policy, as such policies are intended to promote the use of one and only one official language (English, in this case). Because the use of local languages almost always entails at least bilingualism to some degree, so that community members can function at local, regional, and national levels, these languages suffer in countries which are dogmatically monolingual.

National-level attitudes can influence local attitudes. Language is closely associated with the people who speak it; negative attitudes toward a specific

language translate into negative thoughts and beliefs about the speakers and their culture, social norms, and heritage. Such negative views can further influence the views community members have of their language. They may perceive it as backward, useless, underdeveloped, and so on, or they may see it as an impediment to advancement in a larger society which does not value their specific local language. Needless to say, such attitudes have an adverse effect on language use and foster language attrition. The role of community members' attitudes toward their own local language cannot be overstated. Where there is a strong sense of pride in the language, it is more likely to be used and less likely to move into an endangerment situation. In cases where language attrition has begun, the chances of reversing language shift are considerably greater if the people have positive attitudes toward the local language. In the absence of these, a revitalization program must begin by fostering community support.

Causes of Language Shift

The precise causes of language shift are specific to each individual endangerment situation, yet several key factors often come into play. These include urbanization, globalization, and what have been called social dislocation and cultural dislocation. Often the causes of language shift center around imbalances in prestige and power between the local language and culture on the one hand, and the language of wider communication and dominant culture(s) on the other. The imbalance, or unequal levels of power, often means that members of the local community are socially disadvantaged in a number of ways with respect to the majority population. In concrete terms, this frequently means that members of the local community are relatively powerless politically, and are less educated and less wealthy, in many cases living in poverty, and with less access to technology and modern conveniences, than the majority population. One common result is that this socially disadvantaged position becomes associated with, or even equated with, the local language and culture, and so knowledge of the local language is seen as an impediment to social and economic advancement. Socio-economic improvement is thus perceived as tied to knowledge of the language of wider communication, as is renunciation of the local language and culture; for this reason, the situation has been called social dislocation. Social dislocation stemming from lack of prestige and power is one of the most powerful motivating factors in language shift.

Related to social dislocation is what has been called cultural dislocation, which comes about as a result of

modernization and globalization. These two related forces bring people from different cultures, speaking different languages, together in a variety of settings, from informal to official, including religious and educational settings. This often results in the culture of the minority giving way to that of the majority. At an extreme, globalization is feared to lead to cultural homogenization. The loss of cultural distinctions supports a loss of linguistic distinctions, since the culture is embedded in the language.

Urbanization is another key cause of language shift and is itself related to cultural and social dislocation. Urbanization brings people from different regions and cultures into the same living and working spaces. They are necessarily required to communicate with one another and so turn to an established lingua franca or language of wider communication. It is not surprising that we find the highest levels of language retention in rural areas; in general, the more isolated a community, the more likely it is to maintain use of the local language. Urbanization has the opposite effect: by bringing people into contact, it facilitates language shift.

Globalization puts even greater pressure on local languages and can be a major factor in language shift. One of the results of globalization is the emergence of at least one global language of wider communication. A global language is a particular type of language of wider communication, and in some instances may supplant the national language in this role. The global nature of trade and commerce has in recent decades put increasing pressure on the need for an international lingua franca, a position currently held by English. Whereas historically it was important for key figures in world politics to be able to communicate, it is now critical that a large number of people in all walks of manufacturing and business communicate with one another, increasing the need for a global language. Some local communities thus see the knowledge of a global language as necessary for socio-economic advancement. In cases where knowledge of a national or regional language is also important, and in fact may be the only language of education, the need to know the global language can supplant the need or desire to know the local language.

Thus in the modern world, multilingualism generally involves knowledge of one or more national languages and, increasingly, of the global language. This represents a change in traditional patterns, when speakers knew a number of local languages. The shift stems from a combination of factors including education, social prestige, and socio-economics. One factor which has led to diminished local-level multilingualism is the current importance of the national

language in terms of access to education, higher-paying jobs, the media, and social advancement. The national language provides a language of wider communication which makes knowledge of multiple local languages less necessary. A key characteristic of language endangerment is that use of the local language is limited, not only regionally but also functionally. In some cases, it is used only in the home, while in others, it is used in the village but not for communication with people living outside of the immediate community, and so on. Thus, the uses of the local language have become increasingly limited, with the net result that it is increasingly important for speakers to learn not only a language of wider communication but also, in many instances, a global language.

Strengthening Language Vitality

A number of steps can be taken to strengthen language vitality and reverse language shift. These require action and commitment on the part of community members and at the level of national government alike. Such steps include instituting educational programs which teach and promote use of the local language, and establishing national language policies which make these possible and which support linguistic diversity. An often critical part of such programs is the development of literacy in the local language. In most cases, pedagogical materials need to be developed and teachers need to be trained; in cases of advanced attrition, they will need to be taught the language itself, in addition to language pedagogy.

As this implies, levels of extinction and degrees of fluency (especially among semispeakers) are of great relevance to language revitalization efforts. Disappearing languages often have fluent speakers of many ages who can be enlisted in the work of revitalization. For moribund or nearly extinct languages, this is considerably less likely, and the importance of semispeakers to the ultimate success of the process grows considerably. An extinct language may still have rememberers who, although they have no active speaking ability, may know individual words or phrases, such as greetings. So even in cases of extinction there may be a variety of levels of lingering knowledge.

Bibliography

- Bradley D & Bradley M (eds.) (2001). *Language endangerment and language maintenance: an active approach*. London: Curzon Press.
- Brenzinger M (ed.) (1998). *Endangered languages in Africa*. Cologne: Rüdiger Köppe.

- Campbell L & Muntzel M C (1989). 'The structural consequences of language death.' In Dorian N C (ed.), 181–196.
- Crystal D (1997). *English as a global language*. Cambridge: Cambridge University Press.
- Crystal D (2000). *Language death*. Cambridge: Cambridge University Press.
- Dixon R M W (1997). *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Dorian N C (ed.) (1989). *Investigating obsolescence*. Cambridge: Cambridge University Press.
- Fishman J A (1991). *Reversing language shift: theoretical and empirical foundations of assistance to threatened languages*. Clevedon, England: Multilingual Matters.
- Fishman J A (ed.) (2001). *Can threatened languages be saved?* Clevedon, England: Multilingual Matters.
- Grenoble L A & Whaley L J (eds.) (1998). *Endangered languages: current issues and future prospects*. Cambridge: Cambridge University Press.
- Grenoble L A & Whaley L J (2005). *Saving languages: an introduction to language revitalization*. Cambridge: Cambridge University Press.
- Grimes B (2000). *Ethnologue: languages of the world* (14th edn.). Dallas: SIL International.
- Hale K, Craig C, England N, LaVerne J, Krauss M, Watahomigie L & Yamamoto A (1992). 'Endangered languages.' *Language* 68, 1–42.
- Hinton L (2002). *How to keep your language alive*. Berkeley, CA: Heyday.
- Hinton L & Hale K (eds.) (2001). *The green book of language revitalization in practice*. San Diego: Academic Press.
- Janse M & Tol S (eds.) (2003). *Language death and language maintenance: theoretical, practical and descriptive approaches*. Amsterdam: John Benjamins.
- Krauss M (1997). 'The indigenous languages of the North: A report on their present state.' *Senri Ethnological Studies* 44, 1–34.
- Maffi L (ed.) (2001). *On biocultural diversity: linking language, knowledge, and the environment*. Washington, DC: Smithsonian Institution Press.
- Moseley C (ed.) (2004). *Encyclopedia of the world's endangered languages*. London: Curzon Press.
- Mühlhäusler P (1995). *Linguistic ecology: language change and linguistic imperialism in the Pacific rim*. London: Routledge.
- Nettle D & Romaine S (2000). *Vanishing voices*. Oxford: Oxford University Press.
- Robins R H & Uhlenbeck E M (eds.) (1991). *Endangered languages*. Oxford/New York: Berg.
- Skutnabb-Kangas T (2000). *Linguistic genocide in education – or worldwide diversity and human rights?* Mahwah, NJ: Lawrence Erlbaum Associates.
- UNESCO Ad Hoc Expert Group on Endangered Languages (2003). 'Language vitality and endangerment: by way of introduction.' Paris: UNESCO.
- Wurm S (ed.) (2001). *Atlas of the world's languages in danger of disappearing* (2nd edn.). Paris: UNESCO.

Relevant Website

<http://www.ethnologue.com>.

English in the Present Day

M Swan, Didcot, Oxfordshire, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Despite the enormous range of 'Englishes,' there are few important differences between the principal national standards. The following description deals mainly with the common core. Some notes are also included on salient differences between two loosely defined representative varieties, 'general American' and 'standard southern British English.'

Orthography

Modern English spelling has inherited a mixture of Anglo-Saxon, Norman-French, and classical orthographic conventions, many of which were fixed with

the invention of printing in the 15th century. Subsequent phonetic change has been considerable, so that sound–spelling correspondences are now poor. Written English, like French or Irish, is haunted by ghost letters standing for sounds that are no longer pronounced. The 40 or so phonemes of modern English are represented very unsystematically by the 26 letters of the Roman alphabet, singly or in combination. Many phonemes have various possible graphic representations; conversely, many graphemes can represent more than one sound. The situation is particularly chaotic with vowels, where six letters and their groupings serve to encode 17 or more phonemes. As in French, there is a directionality that favors readers over writers: one's chance of guessing how to pronounce a written word is considerably better than one's chance of guessing how to spell a spoken form. It is true that the extent of the orthographic irregularity can be exaggerated – a good deal

of English spelling can be shown to be rule-bound. However, the rules are complex, and many of the most common words are highly irregular and have to be learnt by rote.

There are some superficial American–British differences resulting from Noah Webster’s reforms in the early 19th century. They include:

- AmE *-or*, BrE *-our* (e.g., *color/colour*).
- AmE *-er*, BrE *-re* (e.g., *center/centre*).
- AmE *-og*, BrE *-ogue* (e.g., *catalog/catalogue*).
- AmE *-ize*, BrE *-ise* or *-ise* (e.g., *realize/realise*).
- Different spellings of some individual words. Examples are (AmE first): *aluminum/aluminium*, *analyze/analyse*, *check/cheque* (on a bank), *defense/defence*, *fulfill/fulfil*, *jewelry/jewellery*, *pajamas/pyjamas*, *skillful/skilful*, *tire/tyre* (on a wheel).

More recent moves for spelling reform have failed for predictable reasons: while the ultimate benefits would be substantial, reform would be extremely expensive and disruptive in the short term; and those who would benefit most – children and others who have yet to achieve literacy – are in no position to influence policy. Opponents of reform argue that a reformed orthography could in any case only encode the pronunciation of one variety of English, and would therefore be less than ideal for others; and that reform would obscure lexical relationships such as *photograph*, *photographer* or *sign*, *signature*. Chomsky and Halle (1968) went so far as to claim, indeed, that “Conventional orthography... is a near-optimal system for the lexical representation of English words,” an assertion that might raise eyebrows in the average schoolroom.

Phonology

General

English has a moderately complex phonology: more elaborate than say, Spanish or Japanese, but less so than, for instance, Georgian. There are 24 consonant phonemes (Table 1).

Sixteen of these form voiced/voiceless pairs; though voicing may disappear in word-final position, a fortis/lenis distinction is preserved. Syllable structure permits clusters of up to three consonants (e.g., *strengths* /streŋθs/); clusters of four can occur postvocally in inflected forms (e.g., *glimpsed* /ɡlɪmpst/). /l/, /m/, and /n/, and (in AmE) /r/ can be syllabic.

The vowel inventory lists anything between 14 and 20 phonemes, depending both on the variety being analyzed and the descriptive approach adopted. Typical listings (for notes see ‘American-British Differences’ below) are shown in Table 2.

Table 1 English consonant phonemes

| | |
|------------------|-----------|
| p f θ t s ʃ tʃ k | m n ŋ |
| b v ð d z ʒ dʒ g | r l j w h |

Table 2 American and British vowel phonemes

| Keyword | AmE transcription | BrE transcription |
|---------------|-------------------|-------------------|
| seat | sɪt | sɪt |
| sit | sɪt | sɪt |
| set | set | set |
| sat | sæt | sæt |
| calm, hard | kam, hard | kɑ:m, hæ:d |
| cot, cough | (kət, kɒf) | kɒt, kɒf |
| bought, storm | bɒt, stɔ:m | bɒ:t, stɔ:m |
| book | bʊk | bʊk |
| too | tu | tu: |
| up | ʌp | ʌp |
| turn | (tɜ:n) | tɜ:n |
| ago | ə'go | ə'gəʊ |
| day | de | deɪ |
| my | maɪ | maɪ |
| now | naw | nəʊ |
| no | nə | nəʊ |
| boy | bɔɪ | bɔɪ |
| here | (hɪr) | hɪə |
| there | (ðeɪr) | ðeə |
| tour | (tur) | tʊə |

Vowels are generally oral; nasalization can occur, but is not phonemic. Vowel phonemes differ in intrinsic length, but length is not in itself distinctive.

Prominence, or stress, is realized as a combination of pitch, loudness, and lengthening. Word stress is lexically determined and is not generally predictable from the form of a word (compare *'promise*, *pro'mote*; *'photograph*, *pho'tographer*, *photo'graphic*); initial stress is most frequent. Two levels of stress (in addition to unstress) can be usefully identified (e.g., *enter'tainment*). Some words and multiword items have contextually variable stress (compare *this after'noon*, *'afternoon nap*; *they broke 'up*, *she 'broke up the chair*). Stress is phonemic in a few pairs like *'extract/ex'tract*. Unstressed syllables most frequently have a reduced vowel, usually /ə/ (e.g., *malevolent* /mə'levələnt/), sometimes /ɪ/ (e.g., *decide* /dɪ'saɪd/). Many shorter function words such as articles, auxiliary verbs, prepositions, and pronouns undergo reduction to ‘weak’ forms with reduced vowels in most contexts: compare *I can* /kən/ *hear you* and *Yes, I can* /kən/.

The rhythmic features of spoken English that distinguish it from, say, French or Spanish, arise from a combination of syllable structure, word stress, and vowel reduction; the traditional distinction between

‘stress-timed’ and ‘syllable-timed’ languages is no longer regarded as valid.

Intonation is multifunctional. Raised pitch contributes to the marking of stress. Tone groups mark off syntactic units such as phrases and clauses, with significant falls/rises at boundaries. Like many languages, English signals incompleteness or uncertainty by rising tone movements, and closure or certainty by falling tones. Thus intonation serves not only to distinguish interrogation from assertion, but also to label information as ‘given’ or ‘new’ in discourse. Pitch also contributes to the expression of emotional attitudes.

American–British Differences

While some American and British dialects (e.g., Arkansas and Glasgow) are not mutually comprehensible, speakers of general American and standard southern British English have little difficulty in understanding each other. There are, however, many differences in pronunciation. They include the following:

- BrE has a distinct open back vowel phoneme that is absent in AmE. This is the rounded short /ɔ/ occurring in words like *cot*, *stop*, *on*, *cough*, *often*, *orange*. In AmE these words are pronounced either with /a/ (the same vowel as in *father*, *calm*, *car*) or with /ɔ:/ (the same vowel as in *caught*, *storm*, *all*).
- Standard southern British English is nonrhotic: /r/ is only pronounced before a vowel sound. In general American (as in many other British varieties), /r/ is pronounced in final and preconsonantal positions.
- The disappearance of postvocalic /r/ in BrE has generated new diphthongs in some environments (e.g., *dear* /diə/, *hair* /heə/).
- The AmE intervocalic tapped allophone of /t/, in words like *writer*, does not occur in BrE.
- The glide vowels exemplified in *day*, *no* differ in quality; they are generally analyzed as monophthongs in American transcriptions (/e/, /o/) and as diphthongs in British transcriptions (/eɪ/, /əʊ/).
- The vowel sounds exemplified in *die*, *boy*, and *how* are generally analyzed as phoneme sequences in American transcriptions and as diphthongs in British transcriptions.
- Some words written with *a* + consonant (e.g., *fast*, *after*) have /æ/ in AmE and northern British varieties, and /ɑ:/ in standard southern British English.
- Some words where /u:/ follows a dental or alveolar in AmE have /ju:/ in British English (e.g., *duty*, *tune*, *new*).
- Some words ending in *-ary*, *-ery*, or *-ory* have one more syllable in AmE than in BrE (e.g., *station-(a)ry*).

- French borrowings (e.g. *paté*, *ballet*) tend to have end-stress in AmE and front stress in BrE.

Changes

The last half century has seen much faster changes in pronunciation norms in Britain than in the United States. Up to the 1960s, ‘received pronunciation’ (RP), a nonregional class-based accent used by a small minority of the British population, had the status of a standard. Its prestige, reinforced by its almost universal use in broadcasting, extended to other English speaking countries such as Australia and New Zealand. RP is now spoken by no more than 3% of the population, and to the extent that an influential pronunciation standard continues to exist in Britain, it is a variety closer to vernacular London speech (‘Estuary English’). Vowel quality, in particular, is distinctly different from that of older RP. One consequence of this is that the accent of broadcasters from the 1940s and 1950s sounds decidedly amusing to modern British listeners.

Ongoing changes in pronunciation include the following:

- The use of an unvoiced /ɸ/ or /hw/ in words like *where*, *when* is becoming less common in AmE, and has virtually died out in standard BrE.
- Intrusive /r/ (as in ‘Asiar and Africa’) is on the increase in BrE.
- The change from /ju:/ to /u:/ after dentals and alveolars is continuing, with words like *suit*, *illuminate*, and *enthusiastic* increasingly being pronounced in BrE without /j/, as in AmE.
- Glottalization of medial and final /t/, /p/, and /k/ is growing in BrE.

Lexis

English has an enormous vocabulary – exactly how large is a question to which no clear answer can be given. Well over 600 000 items are included in the *Oxford English Dictionary*, which, however, does not list specialist scientific and technical terminology. In any case, attempts to assess vocabulary size founder on problems of definition. As is becoming increasingly clear from research in phraseology, it is impossible to draw a principled line between orthographic words and other fixed lexical items: compare *cosmetic*, *make-up* (n), and *make up* (v), or *unemployed* and *out of work*.

The different sources of the English word stock are reflected both grammatically and stylistically in the modern language. According to studies cited by Algeo (1998), only 5.4% of the words listed in a dictionary originate from Old English, but these account for

74.5% of the words in newspaper running text. They include most of the highest-frequency vocabulary, and almost all function words such as determiners, auxiliaries, and prepositions. Most other older English words came into the language either from Norman French, or from Latin and Greek during the Renaissance. Besides providing English with much of its learned vocabulary, these later lexical injections have allowed the language to develop a wealth of variant forms for everyday concepts. Speakers and writers can frequently choose between down-to-earth forms (usually, though not always, the older words) and more formal or elevated alternatives (usually the later imports): for example *buy* or *purchase*; *try*, *attempt*, or *endeavor*; *start*, *begin*, or *commence*; *answer*, *reply*, or *respond*; *tell* or *inform*; *get off* or *alight*. For more literary purposes, the possibilities can be almost daunting. Words describing the reflection and transmission of light, for instance, include *shine*, *glitter*, *glisten*, *gleam*, *glow*, *sparkle*, *twinkle*, *glimmer*, *blaze*, and *coruscate*.

The paucity of inflections makes the language morphologically hospitable, and English continues to borrow freely from elsewhere. Algeo (1998) lists 20th-century borrowings from 56 languages, with by far the highest proportion coming from French. Most new English words, however, are home-grown, created either by shifting word class with no formal change (most high-frequency nouns have verbal homonyms, and vice versa), by compounding (e.g., *computer-literate*), or by affixation (e.g., *pro-life*). Over the last century, the latter process has been the major source of new vocabulary. Although English has few inflections, it has a good deal of derivational morphology, with well over 100 affixes in common use, many of them productive. These serve to modify the meanings of words (e.g., *un-*, *counter-*, *re-*, *-ess*, *-hood*, *-ism*, *-ship*), and/or (especially in the case of suffixes) to change their word class (e.g., *en-*, *pro-*, *-age*, *-ance*, *-ful*, *-ly*, *-en*, *-ify*, *-ize*). Among the range of affixes there is a stock of naturalized morphological formative elements, originally borrowed mainly from Greek and Latin, which are particularly productive in present-day word-formation: for example, *auto-*, *eco-*, *cyber-*, *mono-*, *macro-*, *inter-*, *-ology*, *-cratic*, *-phile*, *-phobe*.

While the vast bulk of English vocabulary is common to AmE and BrE, there are a fair number of well-known differences. Many words in common use differ in their reference (e.g., *truck*, *mad*, *pavement*, *chip*); very frequently, different words are used for the same concept (e.g., *elevator/lift*, *faucet/tap*, *check/bill*). The independent development of industry in the two countries is strikingly reflected in, for instance, vocabulary relating to cars

(*hood/bonnet*, *trunk/boot*, *fender/wing*, *gas pedal/accelerator*, *windshield/windscreen*, *gear shift/gear lever*).

Morphology and Syntax

General Characteristics

In traditional typological terms, modern English is situated toward the 'isolating' end of the spectrum. Little of the older inflectional morphology remains; what there is is largely redundant. Nine or so major word classes are commonly distinguished. Grammatical relationships are expressed primarily through word order (SVO) and the use of function words, particularly prepositions and auxiliary verbs. Case structure is nominative-accusative. Topic is not generally distinguished grammatically from subject. Word, phrase, and clause are relatively distinct categories, but there is considerable scope for phrasal and clausal embedding, exploited especially in the formal written language. Constituent order within phrases is mixed, with modifiers generally preceding heads and complements following.

Word Classes

- **Count and mass nouns** are distinguished syntactically; count nouns have an inflected plural form. Formally singular nouns referring to human groups may have partial plural agreement, especially in British English. An inflectional/clitic genitive exists, used preferentially with nouns that have human reference. English differs from most IE languages in having no grammatical gender.
- **Determiners** include a range of quantifiers, demonstratives expressing a two-term distal contrast, and a distributionally and semantically complex article system, similar in most respects to its counterparts in other Western European languages, but using zero article for generic reference (compare German *die Musik*, Italian *la musica*, French *la musique*, English *music*). There is some overlap between determiners and pronouns.
- **Adjectives**, as in most IE languages, form a large and semantically heterogeneous class. Most can function both attributively and predicatively; a few are limited to one or other position. There are semantically based ordering constraints. Comparison is inflectional (with shorter adjectives) or analytic. There is some overlap between the adjective and adverb classes.
- **Pronouns** are personal, reflexive/emphatic, reciprocal, relative, indefinite, possessive, demonstrative, and interrogative. Personal pronouns (first and third person) have distinct case forms; third-person

singular forms encode natural gender. Relative and interrogative pronouns distinguish human from other referents (*who/which; who?/what?*), and retain a moribund case distinction (*who/whom*).

- **Lexical verbs** have a multifunctional base form that serves as a present tense (except in the third person singular), as an imperative, and as an infinitive (usually marked by a particle *to*). There is an inflected third-person singular present, and an inflected past tense. Morphologically marked non-finites are a present participle or gerund, and a 'past participle' identical in regular verbs with the past tense. Irregular verbs form their past tenses and past participles in various ways, mostly involving a vowel change. Participles can combine verbal with adjectival or nominal functions.
- **Primary auxiliaries** are grammaticalized versions of *be* (used to construct passive and progressive verb forms), *have* (used in perfective verb forms), and *do* (used in some interrogative and negative structures).
- **Modal auxiliaries** (*will, shall, would, should, may, might, can, could, must, and ought*) and quasi-modals (e.g., *had better*) express varying degrees and nuances of epistemic, deontic, and dynamic modality, with a good deal of semantic overlap. Modals are invariable, lacking nonfinite and past forms.
- **Adverbs** are a heterogeneous class, including both clausal modifiers and items that modify particular clause elements. Adverb particles such as *back, away, on, off, in, out, over*, up form an important group, overlapping considerably with the preposition class; they combine extensively with verbs to form phrasal verbs.
- **Prepositions and conjunctions** constitute limited but not completely closed overlapping classes, with a growing number of multiword members.
- **Inserts** such as *Hi, Yeah, OK, Sorry, Look, Please, Damn!* form a distinct word class in conversational speech.

Noun Phrase Structure

Nouns can be premodified by determiners, adjectives, and noun modifiers, in that order. Complex hierarchical structuring is possible, facilitated by the freedom with which nouns can act as modifiers; this contributes to the dense and economical approach to information packaging found in some written registers. Some determiners show number agreement; premodifying nouns are generally unmarked for number.

Relative clauses postmodify, as do nonfinite clausal modifiers. Relativized NPs can function as subjects,

direct or oblique objects, or possessives in relative clauses. Restrictive relative clauses have a hybrid character: they can be integrated into a preceding noun phrase by simple juxtaposition, by a relativizer (*that*), or by a relative pronoun (*who, whom, which*), which functions as an argument in the relative clause. The choice is subject to complex syntactic and stylistic constraints.

The Verb Group: Tense, Aspect, Modality, Mood, and Voice

English has a relatively rich and complex verbal system, especially in main clause use. The small inventory of morphological distinctions is supplemented by a range of periphrastic forms constructed with primary and modal auxiliaries. Tense, progressive and perfective aspect, mood, and voice can all be expressed separately or in combination though the verb group. Some key features:

- English perfective aspect encodes anteriority, often with continuing relevance.
We couldn't get in because I had lost my key.
I'm afraid Jane has had an accident.
- Future events arising out of present situations are generally referred to by present forms (especially the present progressive or *be going* + infinitive). Reference to future events viewed as detached from the present uses a bleached modal auxiliary *will*.
We're meeting again tomorrow.
Look out – we're going to crash!
You will be paid at the end of the month.
- 'Irrealis' modality is expressed principally by the use of modal auxiliaries or, especially in subclauses, by past tenses; also (mainly in AmE) by subjunctives.
It would be better if you came tomorrow.
She must have forgotten.
It's important that she be told. (AmE)
It's important that she should be told. (BrE)
- English is unusual in that oblique arguments can act as subjects of passive verbs.
I've been given one of those forms.
She hates being shouted at.
- Reflexive and middle relations can be indicated by pronouns (e.g., *hurt oneself*), but are often unmarked (e.g., *shave, marry*).

Clause and Sentence Structure

Canonical SVO order is generally observed in written English; in informal speech, departures from the norm such as fronting and left-dislocation are more common. Nonprepositional indirect objects precede direct objects. Verb and object constitute a tight unit, not usually separated; otherwise, adverbial elements

occupy a variety of positions. Relative and interrogative pronouns and adverbs are initial in clauses. Preposition-stranding is common. Interrogatives are formed by inversion of the subject and operator (the first of any primary or modal auxiliaries), negation by adding *not* to the operator. (Multiple negation is common in nonstandard dialects.) In the absence of an auxiliary, *do* is used as a dummy operator.

Nominative and accusative case are marked positionally, with redundant morphological marking in some pronouns. There is vestigial subject-verb number agreement in the present tense of lexical verbs and primary auxiliaries. Pronoun dropping is ungrammatical except in imperatives and certain elliptical structures.

Verbs have a wide range of possible complementation patterns; subcategorization rules of individual verbs are only partly predictable on semantic grounds.

Coordination and subordination operate as in most IE languages. Various types of subordinate clause with differing structural features can have nominal, adjectival, or adverbial functions in matrix clauses. Verbal structures in most subordinate clause types exhibit a more reduced range of tense/aspect expression than in main clauses. Embedding can in general be recursive up to the limits of processability. There are complex constraints on extraction from embedded clauses in interrogative and relative structures.

The sentence, as a unit, is essentially a feature of the written language. Conversational speech is not easily analyzed into sentences, and is perhaps better seen as constructed of 'clausal' and 'nonclausal' units (Biber *et al.*, 1999), which may be loosely linked into utterances in an add-on fashion rather than being organized into structural hierarchies.

Information Structure

English follows the quasi-universal tendency to put 'given' or background information before new. Formal written English has few topicalization devices; in order for given information to come first in an utterance without disrupting canonical word order, it therefore tends to be encoded as subject. The merging of topic and subject is facilitated in several ways. English allows an unusually wide variety of participant roles to be encoded as active subjects, including (depending on the verb) agent, patient, experiencer, benefactive, instrument, temporal, and locative. Indirect and prepositional objects can become passive subjects. The 'have-passive' enables experiencer to be encoded as subject (e.g., *She had her roof blown off in a storm*). English is also rich in lexical alternatives with different argument selections (e.g., *admire/impress, frighten/*

fear, notice/strike). In conversational speech, given-new relations are also indicated by fronting, left-dislocation, or phonological prominence.

Changes

Older changes that are still working their way through the language, and some newer developments, include:

- continuing spread of the *going-to* future
- increased use of the progressive with stative verbs (e.g., *I'm understanding German better now*)
- replacement of *shall* by *will*, and of *should* by *would* in some contexts
- spread of the *get*-passive
- decline in the use of the subjunctive, especially in BrE
- general reduction in the use of modals (see Leech, 2003)
- increased use of phrasal verbs
- spread of conditional structures with parallel verb forms in speech (e.g., *If you'd have asked me I'd have told you.*)
- increased use of noun + noun compounds
- disappearance of *whom*
- increased use and acceptability of noncanonical pronoun case in conjoined subjects and objects (e.g., *John and me went ... ; between you and I*)
- grammaticalization of *you guys/folks* as new second-person plural pronoun
- increased use of singular indefinite *they*, partly as alternative to nonsexist *he or she*
- replacement of possessive determiner by object pronoun before *-ing* form (e.g., *This led to them being arrested.*)
- spread of analytic comparatives and superlatives (e.g., *commoner > more common*)
- dropping of complementizer/relative *that*
- increased use of preposition stranding.

For an excellent detailed discussion of ongoing syntactic changes, and further references, see Denison (1998).

American-British Differences

The grammatical systems of standard AmE and BrE are virtually identical. Minor differences, some of which are disappearing with the growing influence of AmE on BrE, include those exemplified below:

| AmE | BrE |
|--|---|
| <i>He just went home.</i> | <i>He's just gone home.</i> |
| <i>Do you have a problem?</i> | <i>Have you (got) a problem?</i> |
| <i>I've never really gotten to know her.</i> | <i>I've never really got to know her.</i> |

| | |
|---|---|
| <i>It's important that he be told.</i> | <i>It's important that he should be told.</i> |
| <i>Yes, I may.</i> | <i>Yes, I may (do).</i> |
| <i>The committee meets tomorrow.</i> | <i>The committee meet/meets tomorrow.</i> |
| <i>He probably has arrived by now.</i> | <i>He has probably arrived by now.</i> |
| <i>It looks like it's going to rain. (informal)</i> | <i>It looks as if it's going to rain.</i> |
| <i>He looked at me real strange. (informal)</i> | <i>He looked at me really strangely.</i> |

Standardization and Prescriptivism

Standardization continues in the English-speaking world, with regional dialects converging on standard varieties, and American English exercising increasing influence on usage in other countries. At the same time, however, increased democratization is reducing the social prestige of the standards, and there is somewhat more tolerance of variation, at least in speech, than, say, 50 years ago. The written-spoken divide is narrowing: with the continued growth of the spoken media, speech is no longer regarded as a poor relation of writing, and characteristically spoken lexical and grammatical elements are making their way into the standard written language. The recent explosion in electronic written correspondence has further reduced the gap, widening the use of relatively informal written registers at the expense of more traditional styles.

However, prescriptive attitudes to usage remain entrenched and powerful in both the United States and Britain. A command of the orthographic, grammatical, and rhetorical conventions of the standard written language is commonly equated with superior intelligence and educational achievement; failure to master these conventions can constitute a serious obstacle to educational or professional advancement. 'Standard' is typically seen as being synonymous with 'correct,' and 'nonstandard' with 'incorrect'; regional or ethnic dialects such as cockney or Afro-American Vernacular English are widely believed to be grammatically deviant and ill-structured.

English is often perceived as being in a state of decay: threatened by variable and changing usage, the encroachment of nonstandard forms, a growing lack of respect for time-honored rules, and the failure of schools to provide adequate grammar teaching. Linguistic and moral decline may be seen as going hand in hand. The British politician Norman Tebbit, echoing voices from the 18th and 19th centuries, famously declared in 1985 that "If you allow standards to slip to the stage where good English is no better than bad English, where people turn up

filthy ... at school ... all those things tend to cause people to have no standards at all, and once you lose standards then there's no imperative to stay out of crime."

The feeling that the language is in danger generates a belief that grammarians, lexicographers, publishing houses, the media, and the educated community have a duty to defend it. In both the United States and Britain, radio and TV stations, magazines, and newspapers receive voluminous correspondence complaining about the 'mistakes' of journalists, broadcasters, and public figures (and, in Britain, condemnation of their 'Americanisms'). Many newspapers and magazines publish regular columns by language gurus on questions of usage. Prescriptive usage guides find a ready market among those who wish to see their prejudices reinforced or who have been made to feel insecure about their command of their own language. (A recurrent advertisement in the British press begins 'Shamed by your mistakes in English?') Such guides often continue to reproduce the 18th- and 19th-century proscriptions against preposition-stranding and split infinitives. Other current targets include novel part-of-speech shifts (e.g., *to task*, *to show-case*); the use of plural concord with group nouns (e.g., *the team are*, standard in BrE, but often condemned for being 'illogical'); 'misplaced' *only*; dangling participles; changes in the meanings of words (e.g., *anticipate*, *disinterested*, *enormity*) and in their grammatical scope (e.g., the use of *hopefully* and *thankfully* as disjuncts); singular indefinite *they* (centuries old); *less* before a plural noun (used by King Alfred); 'misuse' of *me* and *I* in conjoined subjects and objects (there are examples in Jane Austen and Shakespeare); and – perhaps most inflammatory of all – the illegitimate use or omission of the possessive apostrophe. (As Steven Pinker points out [1994], prescriptive rules are often so psychologically unnatural that only those with access to the right schooling can manage to abide by them; so they serve as shibboleths, differentiating the elite from the rabble.)

Despite the growing tolerance of regional pronunciation, elitist attitudes persist: a letter to the *Sunday Times* in March 1993 described speakers of 'Estuary English' as resembling "the dregs of humanity." Particular regionalisms are still widely proscribed or made fun of. British *h*-dropping in words like *horse*, *hurry*, *home*, common in Eastern England, is typically condemned as lazy, slovenly, or slipshod, generally by nonrhotic speakers who do not see their own *r*-dropping at all in the same light.

There have been quite strong counter-currents to prescriptive attitudes in educational and other circles at various times over the last century, but more liberal

and relativistic views of linguistic authority and correctness have tended to meet considerable opposition. Finegan (2001) gives an interesting account of the outrage with which the publication of Webster's 'permissive' *Third New International Dictionary* was greeted in 1961. Jean Aitchison, a British linguist who gave the BBC Reith lecture series in 1996, received copious hate mail for questioning popular views of correctness.

The Future

As English increasingly takes on the role of a world lingua franca, it is certain to develop in interesting ways. One can imagine that a future international variety of English (or group of international varieties) might detach itself somewhat from the American–British standards; alternatively, it might pull them along with it. In either event, narrowly prescriptive attitudes of the kind described above are likely to appear increasingly parochial. While we have no way of knowing how an international English might evolve, we can speculate that it might simplify, regularize or dispense with some of the features that make the present-day language hard for non-native speakers to learn and use. Such changes might include the reduction of consonant clusters; the regularization of word stress and speech rhythm; the simplification of verb phrase grammar, perhaps accompanied by an increase in the range of grammatical particles; the reduction of the modal verb inventory; the regularization of verb complementation structures; and the disappearance of indefinite and

perhaps definite articles. International English might also provide a more favorable environment for spelling reform.

Bibliography

- Algeo J (1998). 'Vocabulary.' In Romaine S (ed.), 57–91.
 Algeo J (ed.) (2001). *The Cambridge history of the English language, volume VI, English in North America*. Cambridge: Cambridge University Press.
 Beal J C (2004). *English in modern times*. London: Arnold.
 Biber D *et al.* (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
 Chomsky N & Halle M (1968). *The sound patterns of English*. New York: Harper and Row.
 Crystal D (1995). *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
 Denison D (1998). 'Syntax.' In Romaine S (ed.), 92–329.
 Finegan E (2001). 'Usage.' In Algeo J (ed.), 358–418.
 Gilman E W (ed.) (1993). *Merriam-Webster's dictionary of English usage*. Springfield, Mass: Merriam-Webster.
 Huddleston R, Pullum G K *et al.* (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
 Leech G (2003). 'Modality on the move: the English modal auxiliaries 1961–1992.' In Facchinetti R, Krug M & Palmer F (eds.) *Modality in contemporary English*. Berlin/New York: Mouton de Gruyter.
 Pinker S (1994). *The language instinct*. London: Allen Lane, The Penguin Press.
 Quirk R *et al.* (1986). *A comprehensive grammar of the English language*. Harlow: Longman.
 Romaine S (ed.) (1998). *The Cambridge history of the English language, volume IV, 1776–1997*. Cambridge: Cambridge University Press.

English, African-American Vernacular

P L Patrick, University of Essex, Colchester, UK

© 2006 Elsevier Ltd. All rights reserved.

The term 'African-American English' (AAE) – formerly 'Negro' or 'Black English' or sometimes 'Ebonics' – denotes a range of ethnically distinctive varieties of North American English characteristically spoken by descendants of Africans brought to the Americas under slavery. This broad range includes regional and social dialects (rural Mississippi folk speech, African Nova Scotian English, Standard AAE); isolated contact and creole varieties (Gullah [Sea Island Creole English], Afro-Seminole Creole in Florida, Texas, and Mexico); distinctive oral discourse styles (preachers' rhetoric, political oratory,

and slang associated with jazz, gospel, or hip-hop); and their literary expressions.

The best known variety of AAE today is African-American vernacular English (AAVE), which is popularly associated with urban culture, working-class and younger speakers, and informal contexts. Contemporary AAVE originated via the Great Migration – the northern urbanization of rural Southern African-Americans that took place between World War I and World War II – and replaced older, post-Emancipation stereotypes of isolated rural dialect speakers, such as the Ex-Slave Elders (Bailey *et al.*, 1991). However, African-Americans collectively speak many varieties of English and other languages: not all use AAVE or participate in vernacular African-American culture (Baugh, 1991, which also surveys

labels for the variety and speakers). Many African-American adults are skilled code-switchers; most have competence in standard AAE, which, although lacking nonstandard grammatical features, is yet distinctively and fluently African-American. Of the 36 million African-Americans in the United States (13% of the population), a large but unknown proportion speaks AAVE.

The distribution and functions of AAE are generally obscured by stigmatization and misrepresentations, principally: mistaking AAVE for all of AAE; the belief that AAVE is not a complete, systematic grammatical system (e.g., mistaking adolescent slang for all of AAVE); and the belief that vernacular use of AAVE is incompatible with literacy and mastery of standard American English (SAE). Linguistic ideologies comprising such beliefs and attitudes have long obscured AAVE's structure and history and have been complicit in racist social structures. Descriptions of AAVE thus generally address not only synchronic structures but diachronic development, attitudes, and applied (especially educational) issues.

Lexicon

African-American speech diversity may be illustrated with the lexicon. Works on AAE list many thousands of items used almost exclusively or with distinctive meaning by African-Americans (Major, 1994; Smitherman, 1994; Cassidy and Hall, 1985). African-American English is a prolific source of slang and specific registers, from which items pass into general American use, if often late and with distortions. Stereotyping and stigmatization of urban slang by non-African-Americans has contributed to the general devaluation of AAVE as a linguistic system. While speaker competence varies along regional, generational, and social lines, knowledge of even widely shared lexical items (unknown in, or contrasting with, mainstream English) does not guarantee familiarity with grammatical properties of AAVE (Green, 2002). The AAE lexicon is a greater unifying force than AAVE grammar. The number of AAE lexical items with clearly identified African etymologies is significantly smaller than in Caribbean English creoles (CECs) – for example, Jamaican Creole English (Cassidy and Le Page, 1967) or Gullah (Turner, 1949), in which hundreds are attested.

Discourse

Uncensored speech (often judged profane or obscene) is subject to different norms in AAVE than in mainstream American English; items for which normalization, neutralization, and generalization have

occurred may differ (Spears, 1998). The stereotypical prominence of uncensored speech, as perceived by non-African-Americans, reflects not only their contemporary social power to censor, but the obsolescence of earlier rules of respectful address and demeanor, enforced on black people by white people. This has resulted in more frequent direct speech (Spears, 2001), including 'dissing', 'reading', and others (Morgan, 1998).

Everywhere in the African diaspora, under slavery, such enforcement spurred development of characteristic modes of indirect speech, ambiguity and speaker agency ('counter-language', Morgan, 1993), e.g., 'signifying', 'marking', and 'loud-talking' (Mitchell-Kernan, 1971). Loud-talking involves direct address and manipulates volume to invoke an audience; marking characterizes targets through direct quotation, which may manipulate features of grammar, phonology, or pragmatics; while conversational signifying attributes "personal characteristics of the target to culturally marked signs" (Morgan, 1998, p. 275). Children and adolescents practice signifying in formalized routines of verbal play known as 'sounding', 'joning', 'snapping', and 'playing the dozens', wittily placing culturally significant emblems (e.g., *yo'mama*) in implausible contexts (Abrahams, 1962; Labov, 1972). African-American English speech events resemble Caribbean ones, like 'busing' (Guyana) or 'dropping remarks' (Barbados), or have pan-African diaspora distribution ('suckteeth', Patrick and Figueroa, 2002).

Syntax

Much early analysis of AAVE grammar contrasted it with both SAE and CECs; recent historical work compares it to Southern White vernacular English (SWVE). Much attention focuses on the AAVE verb phrase, and less on negation, nominal inflection, and pronoun selection (Green, 2002; Rickford, 1999). Major questions include (a) the degree of structural independence of AAVE from general English grammar, (b) evidence for AAVE's historical ancestry and development, and (c) the nature and significance of variation. Grammatical features are more often distinctive, while many phonological features occur in other vernacular English varieties, especially SWVE. Typically, nonstandard features occur more frequently in AAVE, in a wider range of linguistic contexts, while linguistic processes (e.g., rapid-speech reduction rules) are carried further. The nonstandard nature of AAVE is thus gradient, and often exacerbated by comparison to written standard norms.

The AAVE auxiliary system largely mirrors general American English (and so contrasts with CECs).

However, main verbs and auxiliaries (*be*, *have*, *do*) have regular paradigms, without person/number agreement: third singular irregular forms are absent. For auxiliary *be*, the *is* form is generalized to all present persons and numbers (except first singular *am*), and *was* to all in the past. African-American vernacular English carries further a general American English tendency to regularize simple past and present perfect verb forms: they frequently merge in the simple past, although some participial forms are preferred. A few frequent strong verbs use stem forms with past meaning, but past-marking is overwhelming (again, unlike CECs). Future with modal *will* resembles general American English, although the normal phonological reduction to *'ll* is variably extended to complete vocalization and deletion. Question formation optionally involves inverting auxiliaries and modals. Auxiliaries (only) may also fail to surface in questions, and may invert in embedded questions. The AAVE sentential negator *ain't* appears for standard *hasn't/haven't* or *isn't/aren't*, as elsewhere, but also for *doesn't/don't*. As in CECs, the sentential negator is tense-neutral, appearing for past meanings too.

Fundamental contrasts with other dialects occur for AAVE aspectual markers, which take forms deceptively similar to Vernacular American English auxiliaries (*be*, *been*, *done*, *had*). The complex aspectual system makes distinctions unfamiliar to other English dialects. The case of the habitual invariant *be*, widely taken as emblematic of AAVE, is typically misused or misunderstood by nonnative speakers. Other 'camouflaged' forms, often mistaken for their SAE equivalents, include *had* marking the simple past; the nondirectional *come* to express speaker indignation (Spears, 1982); and a stressed *been* for remote past. With stative verbs, *been* denotes continuity to the present moment; stativity constraints are typical of CECs, where *been* also marks remote past. Less familiar AAVE preverbal aspect markers include the immediate future *finna* (<SWVE *fixing to*), the completive *done*, the resultative *be done*, and *steady* marking an intensified continuative. Although several resemble CEC elements in functions, constraints, and even surface forms, few are patterned identically to any CEC, and some came into being or prominence during the last century: habitual *be* was rare "before World War II and . . . virtually non-existent . . . before 1900" (Bailey, 2001, p. 57).

Although negative concord and negative inversion resemble similar strategies elsewhere, AAVE seems unique among the varieties of American English in requiring concord with indefinite object noun phrases, a feature it shares with CECs. As with verbal *-s*, possessive and plural *-s* may both be absent in AAVE, though less often (respectively, 71% absence

compared to 27% and 6% in vernacular Detroit speakers, Wolfram, 1969; absence is significantly greater in CECs). Associative plurals (*Doretha an' (th)em*) resemble CEC plurals with postposed *-dem*, but the AAVE structure (with its mandatory conjunction) occurs in SWVE. Descriptions of generational developments for these complex features are rare (Cukor-Avila, 2001). Pronominal features include pleonastic *it* and *they* in existential constructions. Invariant forms for plural possessives (*yall*, *they*), and occasional object forms for subjects (*us*), represent paradigm levelings that, taken together, are rare to nonexistent in other American English dialects but occur in CECs. The larger issue of modeling AAVE as a (sub-)system *vis-à-vis* general American English remains under-theorized (Labov, 1998).

Phonology

Bailey (2001) surveyed 45 phonological aspects discussed in the literature for either AAVE or SWVE: 19 were shared, 8 were common to AAVE but rare in SWVE (11 were vice-versa), and 6 were pan-English. Several of the oldest established shared features are receding for one or both groups: e.g., loss of inter-syllabic /r/ (*hurry*), and long offglides after /a/ (*half* [hæɪf]). Features that are still vigorous in both varieties, such as glide reduction in /ai/ (*tie*) and /ɔi/ (*boil*), front-stressing (*po'lice*, *De'troit*), or the *pin/pen* merger in [ɛ] before nasals, developed in the late 19th century, when whites and African-Americans worked in contact as tenant farmers. Final consonant cluster reduction, a general English feature, is more frequent in AAVE, significantly before vowels (*firs' apple*).

Features particular to AAVE include shibboleths such as deletion of initial voiced stops /d, g/ in auxiliaries (*I'm (g)onna* /amənə/) and realization of syllable-initial /str/ as [skr] before high front vowels (*street* [skrit]). Others involve final singleton consonants: reduction and loss of nasality in final nasals (*man* [mæ]) and deletion of word-final single consonants after a vowel (*cat* [kæ], *bad* [bæ:]). Final voiced stops are devoiced and sometimes glottalized (*bad* [bæʔ]).

In recent research on vowel subsystems, recordings of mid-19th century speakers show monophthongal /e/ and /o/ and fully back /u/ and /ʊ/, which contrasts with SWVE and American English generally but resembles CECs. Today AAVE possesses the same vowel phonemes as American English; its mergers and glide reductions all occur elsewhere in American English (Bailey and Thomas, 1998). But African-Americans as a group do not participate in systematic shifts of vowel position – the changes from below – that are characteristic of other dialects (Labov, 2001).

Thus /æ/-raising, present in AAVE since the late 19th century, is not integrated into a pattern like the Northern Cities Shift. Several conditioned vowel mergers occur in AAVE and SWVE, but only SWVE reorganizes them into the Southern Shift.

History

The history of AAVE is controversial. Many dialectologists once held an 'Anglicist' position, arguing that AAE showed the same range of variation as comparable white speech (McDavid and McDavid, 1951). This was contested by the early creolist position, which held that a creole developed in colonial America (beyond the Gullah area) and subsequently decreolized, producing current forms of AAVE (Bailey, 1965; Stewart, 1968). Research on variable realization of the *be* copula and auxiliary (Labov, 1969; Baugh, 1980) noted substantial similarities with CECs that were confirmed by later research using methodological refinements (Rickford *et al.*, 1991; Winford, 1992; Blake, 1997).

This line is opposed by a neo-Anglicist position (Poplack, 2000) drawing on data from Ex-Slave recordings and African American congeners (enclaves in Nova Scotia and Samaná), varieties united under the questionable label, 'Early AAE.' Poplack and Tagliamonte (2001) maintain that AAVE is directly descended from British dialects. Analyses of tense/aspect marking compare Early AAE with historical English dialects and CECs and conclude that non-standard features of AAVE "were not created . . . but retained from an older variety of English" (Poplack and Tagliamonte, 2001, p. 251).

Such claims remain controversial among variationists and creolists, whose post-1990 research documents a more complex picture of CEC grammars, which neo-Anglicists largely ignore. The historical significance of enclave varieties rests on their conservative nature and representativeness of their ancestors. Singler's studies (e.g., 1998) of transplanted African-American settlers in Liberia constitute crucial evidence for AAVE; they represent many more speakers (16,000 emigrants) and are more typical of contemporary African-Americans. Singler's results, like Rickford's critiques, frequently conflict with neo-Anglicist positions and emphasize compatibility with recent research on CECs. Wolfram and Thomas (2002), in conducting an intergenerational analysis of an isolated Southern community, also reached different conclusions, arguing that the ethnolinguistic distinctiveness of AAE, although temporarily submerged by accommodation to regional dialect norms, still reflects a broader range of contact influences than British dialects, including African and CEC languages.

While early creolist positions have largely been abandoned in strong form, new research locates creole-like features in 'micro-switches' among indigenous early AAVE speakers far outside the Gullah area (Sutcliffe, 2001). Current creolist positions emphasize multiple causation and structural convergence between CEC and other inputs in a complex contact situation. Speakers from CEC-speaking territories predominated among early black arrivals in many American colonies, often predating direct African imports (Rickford, 1997). Through sustained contact with Gullah, African varieties, and British dialects, CEC speakers contributed substrate influences to AAVE grammar, without leading to acquisition of entire creole grammars (Winford, 1997; Wolfram and Thomas, 2002). Such generalized contact scenarios invoke interlanguage restructuring, for example, in the acquisition of syllable-coda consonant clusters (absent in West African inputs), leading to substantial prevocalic consonant cluster reduction. This produced similar results in both AAVE and CECs (but not SWVE) and thus convergent structures, which might have been reinforced by contact between them (Winford, 1997; Wolfram and Thomas, 2002).

Genesis issues are logically divorced from questions of current change, which focus on convergence or divergence with other American English dialects (Fasold, 1987; Bailey and Maynor, 1989). An expansion of data sources and the range of features studied led to findings of post-Emancipation innovations in AAVE phonology and syntax (examples cited previously). Taken alongside evidence of contemporaneous innovations in SWVE and the discovery of recent rapid diversification among mainstream American English dialects in major cities – in which younger African-Americans, who show surprising agreement on nationwide norms, are not participating (Labov, 2001; Thomas, 2001) – these developments demonstrate considerable divergence. However, convergence on other levels certainly continues (plural *-s* and past-marking, Rickford, 1987), while some features display both (present *-s*, Wolfram, 1987). Moreover, many linguists agree that the evidence is incomplete, the database can be improved, and the complexities of change and variation in AAVE, including its relationships to AAE and American English and the social consequences thereof, are far from satisfactorily understood.

Bibliography

Abrahams R D (1962). 'Playing the dozens.' *Journal of American Folklore* 75, 209–218.

- Bailey B L (1965). 'Toward a new perspective on American Negro dialectology.' *American Speech* 11, 171–177.
- Bailey G (2001). 'The relationship between African American Vernacular English and White vernaculars in the American South: a sociocultural history and some phonological evidence.' In Lanehart S L (ed.). 53–92.
- Bailey G & Maynor N (1989). 'The divergence controversy.' *American Speech* 64, 12–39.
- Bailey G, Maynor N & Cukor-Avila P (eds.) (1991). *The emergence of Black English: texts and commentary*. Amsterdam and Philadelphia: John Benjamins.
- Bailey G & Thomas E (1998). 'Some aspects of African-American vernacular English phonology.' In Mufwene S S *et al.* (eds.). 85–109.
- Baugh J (1980). 'A re-examination of the Black English copula.' In Labov W (ed.) *Locating language in time and space*. New York: Academic Press. 83–106.
- Baugh J (1991). 'The politicization of changing terms of self-reference among American Slave Descendants.' *American Speech* 66, 133–146.
- Blake R (1997). 'Defining the envelope of linguistic variation: The case of "don't-count" forms in the copula analysis of African American Vernacular English.' *Language Variation and Change* 9, 55–80.
- Cassidy F G & Hall J H (eds.) (1985). *Dictionary of American Regional English*. Cambridge, MA: Belknap Press.
- Cassidy F G & Le Page R B (eds.) (1967). *Dictionary of Jamaican English*. New York: Cambridge University Press.
- Cukor-Avila P (2001). 'Co-existing grammars: The relationship between the evolution of African American and Southern White Vernacular English in the South.' In Lanehart S L (ed.). 93–127.
- Fasold R W (ed.) (1987). 'Are Black and White vernaculars diverging?' *American Speech* 62(1). Special Issue.
- Green L J (2002). *African American English: a linguistic introduction*. Cambridge: Cambridge University Press.
- Labov W (1969). 'Contraction, deletion and inherent variability of the English copula.' *Language* 45, 715–762.
- Labov W (1972). *Language in the inner city: studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.
- Labov W (1998). 'Co-existent systems in African-American vernacular English.' In Mufwene S S *et al.* (eds.). 110–153.
- Labov W (2001). *Principles of linguistic change, vol. 2: Social factors*. Oxford: Blackwell.
- Lanehart S L (ed.) (2001). *Sociocultural and historical contexts of African American English*. Amsterdam & Philadelphia: John Benjamins.
- Major C (1994). *Juba to jive: a dictionary of African-American slang*. New York: Penguin Books.
- McDavid R I & McDavid V G (1951). 'The relationship of the speech of Negroes to the speech of whites.' *American Speech* 27, 3–16.
- Mitchell-Kernan C (1971). *Language behavior in a Black urban community*. Berkeley: Monographs of the Language-Behavior Research Laboratory, no. 2.
- Morgan M (1993). 'The Africanness of counterlanguage among Afro-Americans.' In Mufwene S (ed.) *Africanisms in Afro-American language varieties*. Athens: The University of Georgia Press. 423–435.
- Morgan M (1998). 'More than a mood or an attitude: discourse and verbal genres in African-American culture.' In Mufwene S S *et al.* (eds.). 251–281.
- Mufwene S S, Rickford J R, Bailey G & Baugh J (eds.) (1998). *African American English: structure, history, and use*. London and New York: Routledge.
- Patrick P L & Figueroa E (2002). 'Kiss-teeth.' *American Speech* 77(4), 383–397.
- Poplack S (ed.) (2000). *The English history of African American English*. Oxford: Blackwell.
- Poplack S & Tagliamonte S (2001). *African American English in the diaspora*. Oxford: Blackwell.
- Rickford J R (1987). 'Are Black and White vernaculars diverging?' *American Speech* 62, 55–62.
- Rickford J R (1997). 'Prior creolization of AAVE? Socio-historical and textual evidence from the 17th and 18th centuries.' *Journal of Sociolinguistics* 1(3), 315–336.
- Rickford J R (1999). *African American Vernacular English: features, evolution, educational implications*. Oxford: Blackwell.
- Rickford J R, Ball A, Blake R, Jackson R & Martin N (1991). 'Rappin' on the copula coffin: theoretical and methodological issues in the analysis of copula variation in African-American Vernacular English.' *Language Variation and Change* 3, 103–132.
- Singler J V (1998). 'What's not new in AAVE.' *American Speech* 73(3), 227–256.
- Smitherman G (1994). *Black talk: words and phrases from the hood to the amen corner*. New York: Houghton Mifflin.
- Spears A K (1982). 'The semi-auxiliary come in Black English Vernacular.' *Language* 58, 850–872.
- Spears A K (1998). 'African-American language use: ideology and so-called obscenity.' In Mufwene S S *et al.* (eds.). 226–250.
- Spears A K (2001). 'Directness in the use of African American English.' In Lanehart S L (ed.). 239–259.
- Stewart W A (1968). 'Continuity and change in American Negro dialects.' *Florida Foreign Language Reporter* 6(1), 3–4, 14–16, 18.
- Sutcliffe D (2001). 'The voice of the ancestors: new evidence on 19th-century precursors to 20th-century African American English.' In Lanehart S L (ed.). 129–168.
- Thomas E R (2001). *An acoustic analysis of vowel variation in New World English. Publication of the American Dialect Society* 85. Durham: Duke University.
- Turner L D (1949). *Africanisms in the Gullah dialect*. Ann Arbor: University of Michigan Press.
- Winford D (1992). 'Another look at the copula in Black English and Caribbean creoles.' *American Speech* 67(1), 21–60.
- Winford D (1997). 'On the origins of African American Vernacular English – a creolist perspective. Part 1: socio-historical background.' *Diachronica* XIV(2), 305–344.

Wolfram W (1969). *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.

Wolfram W (1987). 'Are black and white vernaculars diverging?' *American Speech* 62(1), 40–48.

Wolfram W & Thomas E R (2002). *The development of African American English*. Oxford: Blackwell.

Relevant Website

<http://privatewww.essex.ac.uk/~patrickp/> – Includes a bibliography of more than 600 items about African-American English.

English, Early Modern

H Raumolin-Brunberg, University of Helsinki, Helsinki, Finland

© 2006 Elsevier Ltd. All rights reserved.

Background

The early modern period (1500–1700) brought several significant changes in the lives of the English people. The most dramatic were perhaps the Reformation, the subsequent dissolution of the monasteries in the 16th century, and the devastating Civil War during the next. Less abrupt phenomena included population growth; changes in the social hierarchy, including greater social mobility; increasing economic activity; a widening world view; and growing national identity.

There were two developments that radically increased the number of written texts. The printing press was introduced into England in 1476, paving the way for subsequent genre diversification and popular writing. The expansion of educational opportunities promoted literacy, although still less than half the men and a third of the women could both read and write, even by 1700.

Good evidence that English had acquired its modern form is that the great literature of this period, such as Shakespeare's plays and the 1611 Bible translation known as the Authorized Version or King James Bible, can be understood by modern readers and listeners despite their archaic feel.

Linguistic Features

Phonology

Understanding Renaissance drama would be more difficult if it were performed using contemporaneous pronunciation, because many Early Modern English (EModE) words were not pronounced the same way as they are today. This divergence largely arises from shifts among the vowels, because changes in the consonant system have been limited.

The well-known series of changes known as the Great Vowel Shift affected long vowels. The origin and course of this development are still in debate, but it is clear that all Middle English (ME) long vowels changed at various times after 1400 (Table 1). The diphthongization of two ME long vowels, /i:/ and /u:/ into /ai/ and /au/, as in *write* and *house*, and the later merger of /e:/ and /ɛ:/, as in *meet* and *meat*, left English with four long vowels, /i:/, e:/, o:/, u:/ . This inventory was supplemented by /ɔ:/, which developed from the diphthong /au/, as in *cause*.

During the early modern period, the ME diphthongs, /iu, eu, au, ai, ou/, were monothongized, becoming /u:/, u:/, ɔ:/, e:/, o:/, whereas /oi/ and /ui/ remained unchanged, as in *joy* and *boil*. A combination of these and new diphthongs that developed from long vowels resulted in three diphthong phonemes, /ai, au, ɔi/, at the end of the period. The impact of postvocalic /r/ which, among other things, often lowered preceding vowels, also created new diphthongs, when an epenthetic [ə] was inserted between a vowel and [r], as in *fire*. Although there may have been sporadic cases of the loss of postvocalic /r/ in the early modern period, its systematic deletion took place later.

Table 1 Change in long vowels: schematic development

| Middle English | Early Modern English (1500–1700) | | Present-day English | |
|----------------|----------------------------------|------|---------------------|-------|
| | Early | Late | | |
| i: | ɔi | ai | ai | write |
| e: | i: | i: | i: | meet |
| ɛ: | ɛ: | e: | i: | meat |
| | | i: | | |
| a: | æ: | e: | ei | make |
| u: | əu | au | au | house |
| o: | u: | u: | u: | food |
| ɔ: | ɔ: | o: | əu | boat |

Sources: Görlach (1991: 70), Barber (1976: 294).

The ME short vowels /i, e, a, o, u, ə/ remained mostly unchanged. A significant change was the southern split of /u/ into two phonemes, /u/ and /ʌ/, e.g., *put* and *cut*.

The consonant system lost the phoneme /x/, which was often spelled <gh>, with its two allophones [x, ç]. Its realizations were replaced by vowel lengthening, e.g., *right* or by /f/, as in *enough*. Two new consonant phonemes, /ʒ, ɲ/, developed, and since that occurred, the inventory of English consonants has remained unchanged. The consonant /h/, was very weak and occasionally dropped, as contemporary spellings like <elmet> for *helmet* and the excrescent <h> in <hevere> for *every* show, but there is no evidence of its stigmatization.

Grammar

Many of the grammatical developments originating in the previous centuries continued in EModE, including morphological simplification and the stabilization of word order. On the whole, English acquired its present analytical structure during the early modern period. These developments are perceptible, for example, in the decline of inflectional endings, the increasing use of auxiliaries, and the diminishment of inversion. Natural gender, in particular, the distinction between human and nonhuman reference, became an important factor among pronouns. The grammaticalization process created new closed-class elements in several areas of grammar.

Nominal inflection was mostly limited to the plural and genitive -s endings, except for some -en plurals, such as *brethren*. The case distinction between the nominative and accusative was only retained in the personal pronouns and the pronoun *who*. The personal pronoun paradigm is indeed a good example of an area that underwent significant changes in EModE (Table 2). First, the second-person plural forms *ye/you/your* came to be used in the singular. This process, which had begun in Middle English, gradually ousted the old *thou* paradigm from Standard English, but allowed subtle expression of power relations, politeness, and intimacy before it did. Second, the originally plural nominative *ye* was replaced by the accusative *you*. Moreover, the first- and second-person possessive determiners *mine* and *thine* lost their -n ending. The last environment where -n was used was the pre-vowel position; e.g., *thine eyes*.

Furthermore, the third-person neuter possessive determiner *its* was introduced into the language around 1600, replacing *his*, which had become affiliated with male human reference as a consequence of natural gender replacing grammatical gender. Human reference also began to constrain the choice of the relative

Table 2 Early Modern English personal pronouns

| | Nominative | Accusative | Genitive (Possessive Determiner) |
|-----------------|------------------------------|------------------|--|
| Singular | | | |
| 1st | I | me | my, mine |
| 2nd | you, ye , <i>thou</i> | you, <i>thee</i> | your, <i>thy</i> , thine |
| 3rd | he | him | his |
| | she | her | her |
| | it, hit | it, hit | its, it , his |
| Plural | | | |
| 1st | we | us | our |
| 2nd | you, ye | you | your |
| 3rd | they | them, hem | their |

Items in bold disappeared during the early modern period. Items in italics became rare during the early modern period. Item in small caps emerged during the early modern period.

pronoun, when *who/whom* was limited to personal reference and *which* to non-personal reference.

Verbal inflection was also simplified by reducing the present indicative suffixes to one, which in most varieties only encoded the third-person singular. The main suffix in the 16th century was -*th*, as in *he maketh*, which gave way to -s in the next century, as in *he makes*. These suffixes were also occasionally used in the plural. The second-person singular suffix -*st* appeared with *thou* but, as mentioned above, this pronoun became rare. The paradigms of strong and irregular verbs were stabilized to a large extent, and the number of verbs with weak conjugation grew.

An important phenomenon in the EModE verb phrase was the increasing use of periphrastic forms and auxiliaries at the expense of one-verb groups. Both *have* and *be* occurred in the (plu)perfect tense, but the use of *be* was limited to mutative intransitives, such as *a childe is come*. The aspectual marker, progressive *be* + *ing*, developed during the early modern era, which also witnessed the development of the auxiliary category, consisting of *can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will*, and *would*, as these verbs lost their full-verb properties, such as non-finite forms and the ability to take non-verbal objects. These auxiliaries were employed to express many of the functions the subjunctive had had in earlier English, although the subjunctive did not entirely disappear.

Furthermore, during the 16th century the non-emphatic use of the auxiliary *do* grew rapidly in affirmative statements, but its popularity fell quickly in the 17th century. In contrast, the employment of *do* in interrogative, imperative, and negative declarative sentences found a firm footing in the language.

Alongside the use of *do* in negative declaratives *dos not hinder*, the old verb + *not* structure was common,

in particular with the verbs *know*, and *doubt*, as in *I know not* and *I doubt not but*. Multiple negation was frequent until around 1600, as in *he helpes me nat with natheng*.

Although the regular word-order in affirmative statements was SVO, in the first decades of the 16th century it was not uncommon to invert the word order after adverbs such as *then*, *therefore*, *thus*, and *yet*. The present-day pattern of inversion after negative adverbs, such as *neither*, *nor*, and *never*, developed from the late 16th century onwards.

New function words were created through grammaticalization. For instance, new indefinite compound pronouns with personal reference developed, such as the forms in *-body*, e.g., *nobody*, from *no* + *body* 'person'. Similarly, complex prepositions such as *because of* from *by* + *cause* + *of* developed in EModE. Several connectives, for instance, *since*, *while*, and *because*, grammaticalized in earlier English, acquired new senses in the early modern period.

Lexis and Semantics

The early modern period witnessed large-scale lexical growth through extensive borrowing and expansion of word-formation patterns. The major source language was Latin, but loans from other languages, in particular French, were also frequent. The extension of English into all areas of life, including science, and the broadening spectrum of genres, styles, and registers created a need for new and more varied vocabulary, which was satisfied with borrowings and native coinages.

This intake of Latinate vocabulary reduced the transparency of the English lexicon by adding a new layer to the already existing Germanic and Romance strata. On the other hand, it provided ample material for the creation of specialist technical terminologies. Not everybody was comfortable with the tremendous increase in Latinate words, as is testified to by the so-called Inkhorn Controversy in the late 16th century, when voices were raised against excessive use of learned borrowings. Dictionary evidence shows that many of these loanwords were short-lived, and it has been estimated that approximately a third of Shakespeare's numerous Latinate neologisms only occurred once.

Loanwords became integrated into the language to varying degrees. Writers occasionally used nearly synonymous word-pairs to make sure that the readers understood the borrowing, e.g., *wepying and lamen-tyng*. Inflectional endings such as L *-atus* > E *-ate*, were modified as in *alternate*. Multiple borrowing occurred, and the spelling of older French loanwords could be Latinized, as illustrated by *dout* > *doubt*.

The existing word-formation patterns, affixation, compounding, and conversion, were expanded by borrowing. For instance, among the negative prefixes, the native *un-*, e.g., *unmeet*, encountered four competitors, *a-*, *in-*, *dis-*, and *non-*, e.g., *discontented*.

The biggest problems for understanding EModE texts arise from semantic changes, which range from radical changes in meaning to shifts in nuances. For example, *mete* means 'food' in general and not what we understand by *meat* today, and *an auncient and sad matron* in Sir Thomas Elyot's educational treatise is simply 'an old and trustworthy married woman'.

Linguistic Diversity

Although the presentation of early modern general features may give the impression of linguistic uniformity, we must not forget that there was substantial regional, social, individual, generic, and stylistic variation. Despite the lack of a codified model, the language underwent standardization in terms of focusing and option-cutting, with orthography leading the process. The spellings of early and later texts diverge in general, but barely literate people's private writings contain atypical spellings at all times.

There is no doubt that people spoke their local dialects, but dialectal texts are rare because of the low level of literacy. Drama can offer some information on dialectal usage, and private texts occasionally contain regional features.

As in any society, there were social differences in the language use in early modern England, which was markedly hierarchical. Research on the diffusion of morphosyntactic changes shows that several were introduced by the middling ranks and diffused from the London region to the rest of the country. Changes

Table 3 The Early Modern English genres in the *Helsinki Corpus* of English texts

| | |
|-----------------|-----------------------|
| Literate genres | Law |
| | Handbooks |
| | Science |
| | Educational treatises |
| | Philosophy |
| | Sermons |
| | History |
| | Travelogue |
| | Biography |
| | Letters non-private |
| | Bible |
| | Trial proceedings |
| | Fiction |
| | Letters private |
| Oral genres | Drama |
| | Diary |
| | Autobiography |

originating in the spoken language were usually led by women, but men were ahead of women in learned changes such as the loss of the multiple negation. The diversification of genres led to new genre conventions and stylistic variation. The 17 early modern genres included in the *Helsinki Corpus of English Texts* illustrate the broad spectrum of early modern writing (see Table 3). The most formal literate genres such as law, philosophy, and science stand out with Latinate diction and structural complexity both on the sentence and phrase level, whereas oral genres such as fiction and private correspondence often rely on Germanic words and simpler structures.

Bibliography

- Barber C (1976). *Early Modern English*. London: Andre Deutsch.
- Görlach M (1991). *Introduction to Early Modern English*. Cambridge: Cambridge University Press.
- Kastovsky D (ed.) (1994). *Studies in Early Modern English*. Berlin & New York: Mouton de Gruyter.
- Lass R (1999). 'Phonology and morphology.' In Lass R (ed.). 56–186.
- Lass R (ed.) *The Cambridge history of the English language*. Vol. III, 1476–1776. Cambridge: Cambridge University Press.
- Michigan Early Modern English Materials*. Compiled by Richard W Bailey, Jay L Robinson, James W Downer & Patricia V Lehman. University of Michigan.
- Moessner L (ed.) (2001). 'Early Modern English text types.' Special issue of *European Journal of English Studies* 5/2.
- Nevalainen T (1999). 'Lexis and semantics.' In Lass R (ed.). 332–458.
- Nevalainen T (2005). *An introduction to Early Modern English*. Edinburgh: Edinburgh University Press.
- Nevalainen T & Raumolin-Brunberg H (2003). *Historical sociolinguistics: language change in Tudor and Stuart England*. Longman Linguistics Library. London: Pearson Education.
- Rissanen M (1999). 'Syntax.' In Lass R (ed.). 187–331.
- Rydén M, Tieken-Boon van Ostade I & Kytö M (eds.) (1998). *A reader in Early Modern English*. Bamberger Beiträge zur Englischen Sprachwissenschaft 43. Frankfurt a/M: Peter Lang.
- Salmon V & Burness E (eds.) (1987). *A reader in the language of Shakespearean Drama*. Amsterdam & Philadelphia: Benjamins.
- Schendl H, Dollinger S & Hirsch A (eds.) (forthcoming). *A comprehensive bibliography of linguistic writings on Early Modern English*. CD-ROM. Berlin: Mouton de Gruyter.
- Stein D & Tieken-Boon van Ostade I (eds.) (1994). *Towards a Standard English, 1600–1800*. Berlin & New York: Mouton de Gruyter.
- The Corpus of Early English Correspondence (CEEC)*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi and Minna Palander-Collin. University of Helsinki.
- The Corpus of Early English Correspondence Sampler (CEECS)*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi & Minna Palander-Collin. University of Helsinki: OTA & ICAME.
- The Helsinki Corpus of English Texts (HC)*. Compiled by the Helsinki Corpus project team, Matti Rissanen as director and Merja Kytö as secretary. The Early Modern English section mainly compiled by Terttu Nevalainen & Helena Raumolin-Brunberg. University of Helsinki. OTA & ICAME.
- The Helsinki Corpus of Older Scots (HCOS)*. Compiled by Anneli Meurman-Solin. University of Helsinki. OTA & ICAME.
- The Lampeter Corpus of Early Modern English Tracts*. Compiled by Josef J. Schmied, Claudia Claridge & Rainer Siemund. Chemnitz University of Technology. OTA & ICAME.
- The Parsed Corpus of Early English Correspondence (PCEEC)*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi & Minna Palander-Collin. Tagged by Arja Nurmi, University of Helsinki, and parsed by Ann Taylor, University of York. OTA & ICAME.
- The Penn-Helsinki Parsed Corpus of Early Modern English*. The Helsinki Corpus Early Modern English section extended, tagged and parsed by Anthony Kroch, Beatrice Santorini and Lauren Delfs. University of Pennsylvania. <http://www.ling.upenn/emodeng>.

Related Websites

- <http://ota.ahds.ac.uk> – Oxford Text Archive (OTA), Oxford. The OTA has a large number of early modern texts freely available for non-commercial use.
- <http://www.ling.upenn/emodeng>
- <http://helmer.aksis.uib.no/icame> – The International Computer Archive of Modern and Medieval English (ICAME) at University of Bergen, Norway.
- <http://chadwyck.co.uk> – Chadwyck-Healey, Cambridge, UK, provides additional corpora on Early Modern English literature and printed books.

English, Later Modern (ca. 1700–1900)

J Beal, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The recognition of Late Modern English as a separate period in the history of the English language is a fairly recent phenomenon. The first linguist to use the phrase appears to be Poutsma, but his (1914) *A grammar of Late Modern English* was effectively a synchronic study of what, to him, was present-day English. Poutsma describes his work as “a methodical description of the English Language as it presents itself in the printed documents of the last few generations” (Poutsma, 1928: viii). Sweet seems to have invented the now familiar tripartite division of the history of English when he proposed in a lecture to the Philological Society to “start with the three main divisions of *Old*, *Middle* and *Modern*, based mainly on the inflectional characteristics of each stage” (Sweet, 1873–1874: 620). His definition of Modern English was the period of lost inflections, i.e., from the 16th century to Sweet’s own lifetime. It was not until the early 20th century that a subdivision of Modern English was called for by Wyld:

We should further distinguish Early Modern, from 1400 or so to the middle of the 16th century; and after that it is convenient to distinguish late 16th-century, 17th-century, and 18th-century English and we may consider present-day English to begin toward the end of the 18th century (Wyld, 1936: 27).

Wyld thus recognized a distinction between Early Modern English and the later history of the language, but does not see this later period as a coherent entity worthy of a single label. As I have pointed out elsewhere (Beal, 1999, 2004), the study of Late, or Later Modern English, as it is sometimes termed, has been the ‘Cinderella’ of English historical linguistics: serious study of this period did not begin until the end of the 20th century, perhaps because it was not until the millennium was in sight that the Late Modern period could be seen in historical perspective. This is the explanation provided by Charles Jones:

There has always been a suggestion . . . especially among those scholars writing in the first half of the twentieth century, that phonological and syntactic change is only properly observable at a great distance and that somehow the eighteenth, and especially the nineteenth centuries, are ‘too close’ chronologically for any meaningful observations concerning language change to be made (Jones, 1989: 279).

Jones has been a pioneer in the study of Late Modern English, organizing the first conference dedicated to this period in 2001, the proceedings of which have appeared as Dossena and Jones (eds.) (2003). Görlach has produced separate volumes dedicated to 18th and 19th century English, respectively (Görlach, 1999, 2001), and Bailey (1996) likewise deals with the 19th century separately, leaving Beal (2004) as the first book to cover the whole of this period in a single volume. Although the latter defines ‘later’ Modern English as covering the period 1700–1945, I will deal here with the period (roughly) from 1700 to 1900, leaving the 20th century to the article **English in the Present Day**.

External History

Although the division between the Early and Late Modern periods is not defined by a single cataclysmic external event such as the Norman Conquest, there are several factors contributing to the view that a date around 1700 marks the beginning of a new era. Historians generally describe the ‘long’ 18th century as beginning with the Restoration of the monarchy in England (1660) and ending with the fall of Napoleon (1815), while the ‘long’ 19th century stretches from the start of the French Revolution (1789) to the end of World War I (1918). These overlapping ‘long’ centuries together provide a good working definition of the Late Modern period in English. ‘Restoration’ is perhaps a misnomer for the accession of Charles II, since, far from occupying the throne by Divine Right, he had been invited to do so by Parliament. The so-called Glorious Revolution of 1688 and the Bill of Rights (1689) further reduced the powers of the monarch, bringing in the constitutional system that still exists in Britain.

1660 is also the year when the Royal Society was founded, ushering in the Age of Reason. Both Porter (2000) and Lass (1999) cite the publication of Newton’s *Principia* (1687) as the beginning of the English Enlightenment. Scientific progress from throughout the Late Modern period stimulated lexical innovation as new inventions, processes, and whole disciplines required names. The emphasis on ‘reason’ in the 18th century also contributed to the climate of opinion, described by Leonard (1929) as the Doctrine of Correctness, in which grammatical ‘rules’ such as the proscription on negative concord (‘two negatives make a positive’) were rationalized on mathematical models. The scientific discoveries of the late 17th and early 18th centuries led to the technological innovations that drove the Industrial

Revolution of the late 18th and 19th centuries. As Britain became an industrial nation, workers moved from the countryside to the newly expanding towns and cities, especially in the English North and Midlands. In cities such as Manchester, Leeds, and Birmingham, new urban dialects evolved as the demand for labor brought in workers from various parts of Britain and beyond.

Dialect contact, and an awareness of the differences between dialects of English, was made possible by advances in transport and communications during the Late Modern period. In the course of the 18th century, the Turnpike Trusts funded a substantial number of new roads, cutting the journey from York to London from three days to one, thus making leisure travel a more pleasant and practical proposition. Travelogues such as Defoe's *A tour thro' the whole island of Great Britain* (1724–1727), Johnson's *Journey to the Western Isles of Scotland* (1755) and Cobbett's *Rural rides* (1830) describe journeys by carriage undertaken for curiosity. The 'outlandish' dialects of Northumberland and Cornwall (the areas of England most remote from London) are as much objects of curiosity for Defoe as the landscape and customs of other areas. In the course of the 19th century, the development of the railway made affordable leisure travel possible for the lower and middle classes, at least in urban areas. Communication was further facilitated by the introduction of the Penny Post in 1840 and the electric telegraph in 1837, while the invention of the phonograph in 1877 made it possible to hear the disembodied voices of speakers from distant places. All these developments had the effect of increasing dialect contact between speakers of different British dialects and, eventually, between speakers (and writers) of British and American English.

The Late Modern period also marks the beginning of the 'great divide' between British and American English. Although the first English-speaking colonies in what is now the United States were founded in the early 17th century, the development of American English as a national variety with its own prescribed norms was precipitated by the American Revolution (1775–1783). In 1789, Webster asserted that 'customs, habit and *language*, as well as government, should be national. America should have her *own*, distinct from all the world.' (Webster, 1789: 179). His *American dictionary of the English Language* (1828) provided norms for spelling which were deliberately differentiated from those of British English, as well as legitimizing Americanisms. With the loss of America, British colonial expansion diverted to Australia (1788) and, in the 19th century, to South Africa and New Zealand. The development of distinct national

varieties of English in these countries was perhaps more a phenomenon of the 20th century, but the expansion of British interests both in these colonies and the nations of Africa and Asia absorbed by the British Empire, brought into English loan words from a wide variety of languages, as flora, fauna, topographic features, and customs hitherto unknown to speakers of English required names. This, along with the scientific discoveries and inventions referred to above, accounts for the increase in lexical innovation during the 19th century.

Within Britain, increased social mobility brought about by the commercial opportunities of the Industrial Revolution and the expansion of educational provision in the course of the Late Modern period led to the emergence of an ambitious and influential middle class. Such 'social climbers' created a market for the normative texts for which this era is famous: alongside the 'triumvirate' of Johnson (1755), Lowth (1762), and Walker (1791), many other dictionaries, grammars, and pronouncing dictionaries were published to help those aspiring to linguistic correctness. Whether such normative works had any effect on the language, or whether they simply described linguistic practices already in use among the educated, is still a matter of debate (see, for instance, Beal, 2003), but it is certainly the case that the codification of Standard (British) English is one of the defining features of the Late Modern period. The following sections provide a brief account of the major changes that took place in the Late Modern English Period, but it is to be understood that this refers only to Standard English in England. For discussion of the history of other varieties of English, see Beal (2004: 190–220) and Watts and Trudgill (2002).

Morphology and Syntax

Morphology

We have seen above that Sweet defined the Modern English period as a whole as the period of lost inflections. However, this definition refers largely to the Early Modern period, as the only inflection to be lost after 1700 is the second person singular-*st* ending. This in turn depends on the loss, from Standard English, of the distinction between second person singular *thou*, *thee*, *thy*, *thine* and the formerly plural *ye*, *you*, *your*, *yours*. The singular forms had become marked in the Early Modern period, and by 1700 "survived only in dialects, among Quakers, in literary styles, as a device of heightening . . . and in its present religious function" (Strang, 1970: 140). Although some 18th-century authors, notably Sheridan and Richardson, put *thou* forms into the mouths of

upper-class males, this usage was not universally accepted. McKnight (1928, reprinted 1968: 335) cites Greenwood (1711) as stating “it is counted ungentile and rude to say, *Thou dost so and so*.” Of course, the loss of *thee/ thou* pronouns and the *-st* verb inflection (along with *art, wert* forms of *be*) meant that Standard English no longer marked the distinction between second person singular and plural. In the early 18th century, the distinction was maintained, at least with *be*, by using *you was* for the singular and *you were* for the plural. However, Lowth condemned this as an “enormous solecism” (Lowth, 1762: 48), and *you was* disappeared from Standard English, though it is still common in dialects. By the end of the 19th century, a new plural form *yous* was being used in Irish, American, and Australian English (Wright, 1898–1905), and this form has spread throughout Britain in the late 20th century (Cheshire *et al.*, 1993).

Syntax: Regulation of Early Modern Variants

Syntactic changes in the Late Modern period can be divided into two types: those that represent the final stages of processes begun in the Early Modern period, and those that are innovations of the 18th and 19th centuries. The former type involve the regulation of variants, a process to some extent helped along by the prescriptive grammarians of the Late Modern Period. Changes in the auxiliary system led, in the Early Modern period, to variation between positive declarative clauses with and without *do*. While the 16th-century grammarian Palsgrave was able to state: “*I do* is a verb moche comenly used in our tonge to be put before other verbs, as it is all one to say ‘*I do speake*’, and such lyke, and ‘*I speake*’” (Palsgrave, 1530: 523), Johnson condemned this usage as “a vitious mode of speech” (Johnson, 1755: Sig. B2v). Ellegård (1953: 162) provides a graph demonstrating the decline of this construction through the Early Modern period, reaching close to zero by 1700. However, Tieken (1987) found that, in a corpus of 18th-century prose (948,700 words), the construction was still used, but it was rare, and 13 out of the 14 examples she found occurred before 1760. In poetry, the use of *do* before another verb persisted into the 19th century, presumably because the semantically empty auxiliary here provided an extra syllable when needed, and allows the citation form of the verb to be placed at the end of a line, thus facilitating rhymes. An example from Wordsworth (1827) is:

The hapless creature which did dwell
Erewhile within the dancing shell.
(*The blind highland boy* 193–194)

However, it is certainly the case that this usage was extremely marked after the mid-18th century, at least in Standard English. Another, related, area in which Early Modern variants persist into the 18th century is negation: Ellegård (1953: 162) shows the use of *do* in negative declaratives in his corpus rising to 80% by 1700, with the steepest rise in this usage coming in the second half of the 17th century. Tieken (1987) found that an average of 76% of the negative declaratives in her 18th-century corpus involved the use of *do*, and that the decline of the *do*-less negative was gradual throughout the 18th century. She also found that the *do* + *not* + infinitive construction was more frequent in the usage of more educated authors, but that the *do*-less form was most frequent with the verbs *know* and *doubt*.

Another area in which regulation took place during the Late Modern period is relativization. The so-called *wh*-relatives *who*, *whom*, *whose*, and *which* had been introduced in the Early Modern period, but the options of using *that* with both human and nonhuman antecedents, and of using a zero, or contact relative in both subject and object positions, were still available. The 18th century saw grammarians expressing a preference for *who/ whom/ whose* with human antecedents and *which* with nonhuman antecedents, and condemning the zero relative. However, although Visser (1963–1973: 540) states that “a remarkable decline in the currency of the zero-construction becomes perceptible” in the 18th and 19th centuries and notes that Johnson called this “a colloquial barbarism,” Raybould (1998) notes that Johnson uses this construction even in the nominative. It would appear that 18th-century authors saw this construction as informal or colloquial rather than strictly ungrammatical. The only substantive changes in relativization in Late Modern English are the restriction of *which* to use with inanimate antecedents, and the restriction of zero-relatives in the nominative to colloquial usage, and to existential constructions such as *there’s a man out here wants to see you*.

Syntax: Innovations of the Late Modern Period

Perhaps the most fully researched feature of Late Modern syntax is what has been called the ‘*be* + *-ing*’ construction, the ‘progressive’ and the ‘expanded form,’ as in *She is reading a book*. Although this construction had been used before 1700, its use in Early Modern English was optional in contexts where today it would be required. Thus in *Hamlet* II. 2. 190, Polonius asks Hamlet “What do you read my Lord?” Today this would be interpreted as an inquiry into Hamlet’s reading habits, but Polonius was referring

to the book that Hamlet had in his hands at the time. Today, the required construction in this pragmatic context would be *What are you reading, my Lord?*

Several studies (Dennis, 1940; Strang, 1982; Arnaud, 1983) have noted the increase in usage of the *be* + *-ing* construction throughout the Late Modern period, both in terms of the sheer numbers of such constructions, and the types of clause in which it can occur. Strang notes that before 1750 the construction is used mainly in subordinate clauses, but that, from the second half of the 18th century, the rise in the use of *be* + *-ing* constructions is proportionately greater in nonsubordinate clauses. Thus, the construction becomes fully grammaticalized in the course of the Late Modern period. Likewise, from the second half of the 18th century, there is a rise in the use of *be* + *-ing* with stative verbs such as *love*, *wish*, etc., with verbs denoting instant actions such as *explode*, *fall*, etc., and with nominal and adjectival complements (e.g., *You're being a fool/foolish*). The extension of *be* + *-ing* to the passive is likewise an innovation of the Late Modern period. As late as 1870, this construction was still being condemned by grammarians such as Marsh, who described it as “at war with the genius of the English Language” (Marsh, 1860: 465). The preferred form, for Marsh, was *the house is building*. However, this had in turn been condemned by Johnson (1755) as an unacceptable innovation:

The grammar is now printing, brass is forging ... in my opinion a vitious expression probably corrupted from a phrase more pure but now somewhat obsolete: a printing, a forging.

While examples of the *a-printing* type are found in early 18th-century literature, by the 19th century they are being used (*sic.*) to represent nonstandard speech. In Richardson's *Pamela* (1740), the sentence “this girl is always a-scribbling” is given to an educated, upper-class man, but in George Eliot's *The mill on the floss* (1860), the same construction conveys uneducated, lower-class usage: “I hope, sir, you're not a-thinking as I bear you any ill-will ... I'm not a-defending him.” The first examples of the passive with *be* + *-ing* are found in late 18th-century letters, such as the following, cited in Denison (1998: 152):

I have received the speech and address of the House of Lords; probably, that of the House of Commons was being debated when the post went out. (1772, Harris, *Letters*)

Letters are, of course, the most informal genre of writing: we have already seen that Marsh was reluctant to accept this construction a century later, and,

even in the early 20th century Curme and Kurath seem to begrudge it:

From 1825 on ... the form with *being* + perfect participle began to lead all others in this competition, so that in spite of considerable opposition the clumsy *is being built* became more common than *is building* in the usual passive meaning, i.e. where it was desired to represent a person or thing as affected by an agent working under resistance vigorously and consciously to a definite end: ‘The house *is being built*’, ‘My auto *is being repaired*’ (Curme and Kurath, 1931: 444)

The extension of the use of the *be* + *-ing* construction to longer verb phrases involving perfective and modal verbs took longer. Marsh constructs such sentences as artificial examples of what he sees as the ludicrous consequences of allowing the passive with *be* + *-ing*:

They must say therefore ... the great Victoria bridge *has been being built* more than two years; when I reach London, the ship Leviathan *will be being built*; if my orders had been followed, the coat *would have been being made* yesterday; if the house had then *been being built*, the mortar would *have been being mixed* (Marsh, 1860: 654).

However, Marsh's ridicule was in vain, for, by the early 20th century, such constructions were in use: they were rare then as now, simply because the pragmatic circumstances in which they might be used are likewise rare. An early example is from Galsworthy: “She doesn't trust us: I shall always be being pushed away from him by her” (Galsworthy, 1915, *Freelands*, cited in Denison, 1998: 158).

The extension of the *be* + *-ing* form to the passive and to other paradigms is perhaps the most significant syntactic innovation of the Late Modern period. Other changes tend to involve, as Denison notes, “a given construction occurring throughout the period and either becoming more or less common generally or in particular registers” (Denison, 1998: 93). The role of prescriptive grammarians in relegating constructions such as the double negative to nonstandard usage in this period is a matter of debate. Greenwood's statement that “two *Negatives* or two *Adverbs of Denying*, do in *English* affirm” (Greenwood, 1711: 160) is often cited as an example of mathematical logic inappropriately applied to language. Yet Tieten (1982) and Austin (1984) note that, in the 18th century, multiple negation occurs in informal and lower-class writing, and in the portrayal of such usage by playwrights. Whether the grammarians created the stigma or merely reflected the sociolinguistic situation of their day is as difficult a question as that of prioritizing the chicken and the egg. What is certain is that multiple negation disappears from formal Standard

English in the course of the Late Modern Period. Other constructions condemned by grammarians, such as preposition stranding and the split infinitive remain shibboleths to this day, and are perhaps more common in informal Standard English, but have by no means disappeared. In fact, Lowth, who is often cited as proscribing preposition stranding, merely states that this “is an idiom, **which our language is strongly inclined to**” and that this “prevails in common custom, and suits very well with the more familiar style in writing; but the placing of the Preposition before the Relative is more graceful . . . and agrees much better with the solemn and elevated style” (Lowth, 1762: 167–168, my emphasis). As Tieken (2000) points out, Lowth’s use of the very construction he is supposedly condemning, is intended as a joke, and, far from proscribing preposition stranding, Lowth states that it is perfectly suitable for informal writing and, indeed, uses it in his own letters. His statement appears to be a description of 18th-century usage. The split infinitive was never mentioned by Lowth or any other 18th-century grammarian: it is first mentioned in 1834 in the *New England Magazine*, where it is represented as “not unfrequent among uneducated persons” (Lowth, 1834: 469, cited in Bailey, 1996: 248). If educated persons avoid the split infinitive in formal writing today, this is largely because it has become such a shibboleth.

Phonology

The phonology of Late Modern English has, until very recently, had much less scholarly attention paid to it than that of earlier periods. This is probably because, as MacMahon suggests “superficially, the period under consideration might appear to contain little of phonetic and phonological interest, compared with, for example, earlier changes such as the transition from Old to Middle English, and the Great Vowel Shift” (MacMahon, 1998: 373). It is in the Late Modern period that, as Holmberg so neatly puts it, “the snob value of a good pronunciation began to be recognised” (Holmberg, 1964: 20), and elocutionists such as Thomas Sheridan and John Walker made good livings from providing lectures and pronouncing dictionaries (Sheridan, 1780; Walker, 1791) to the upwardly mobile. This is also the period in which Received Pronunciation emerged as the sociolect of the public-school-educated aristocracy and upper-middle class, eventually to become the reference variety of British English. When we discuss changes in English phonology and phonetics in this period, it has to be understood that we are comparing the reference varieties of the 18th and 19th centuries, as set out, for instance, in successive editions of

Walker’s *Critical pronouncing dictionary* from 1791 to 1904, with 20th-century RP as defined in those of Daniel Jones (first edition 1917). Other varieties of present-day English tend to retain variants ‘left behind’ by Late Modern English sound changes such that a passage transcribed according to Walker (1791) would sound regional and/or slightly archaic rather than obsolete or outlandish to a 21st-century British ear (see Beal, 2004: 134 for such a transcription). Although elocutionists such as Walker and Sheridan were undoubtedly normative, their detailed descriptions of sounds and transcriptions of every word in their dictionaries provide a great deal of evidence for the prestigious pronunciation of the period. A more detailed account can be found in Beal (1999 and 2004: 125–167): here, I will briefly describe the main changes in the pronunciation of received English between 1700 and 1900.

Consonants

Perhaps the most striking difference between the pronunciation set out in Walker (1791) and present-day RP is that the former is rhotic, with orthographic <r> pronounced in all positions. The loss of rhoticity is attested by Walker, who is one of the first sources of evidence for this change, but he considers this to be a marker of lower-class London usage:

In England, and particularly in London, the r in *lard*, *bard*, *card*, *regard*, is pronounced so much in the throat, as to be little more than the middle or Italian a lengthened into *baa*, *baad*, *caad*, *regaad* (Walker, 1791: 50).

He goes on to describe the Irish pronunciation of /r/ as too harsh, but to say that the pronunciation at the beginning of a word should be more ‘forcible’ than at the end, so that ‘*Rome*, *river*, *rage*, may have the r as forcible as in Ireland, but *bar*, *bard*, *card*, *hard*, etc. must have it nearly as soft as in London (Walker, 1791: 50). This suggests that, even in the pronunciation recommended by Walker, /r/ was considerably weakened in final and preconsonantal positions. However, as Mugglestone (1995: 98–103) demonstrates, ‘dropping’ of <r> continued to be overtly stigmatized until the late 19th century. Loss of rhoticity in British, or, rather, English English, appears to have been a ‘change from below,’ first noticed in lower-class London English of the late 18th century and eventually to find its way into RP. In the 20th century, rhoticity became recessive even in regional dialects of England, remaining as a marker mainly of southwestern and some Lancashire dialects.

The other main consonantal changes in Late Modern English are not so much changes in the system, or even the distribution of phonemes, as the regulation of variants. Two of the greatest shibboleths

of nonstandard pronunciation in the 20th and 21st centuries are popularly known as ‘dropping’ of <h> and <g>. In the latter case, the term ‘dropping’ is not at all accurate, since the stigmatized variant is /n/ as opposed to /ŋ/ in, for example, *hunting*, *shooting*, and *fishing*. In both cases, the stigmatized variants had been attested at least from the Early Modern period, but are not labeled as ‘vulgar’ or ‘incorrect’ before the 18th century. Sheridan was the first to comment on ‘h-dropping’:

There is one defect which more generally prevails in the counties than any other, and indeed is gaining ground among the politer part of the world, I mean the omission of the aspirate in many words by some, and in most by others (Sheridan, 1762: 34).

Walker lists among the ‘faults of the Cockneys’ that of “not sounding *h* where it ought to be sounded, and inversely” (Walker, 1791: xii–xiii). For both Walker and Sheridan, the dropping of /h/ from the /hw/ of *which*, *what*, etc. is as much a ‘fault’ as the omission of initial /h/ in *house*, etc. While h-dropping remains highly stigmatized in all but a handful of words of French origin (*hour*, *honour*, *herb* and derivatives), the initial /hw/ is now marked in RP as very conservative. For a full account of the extent to which h-dropping became a shibboleth in the course of the Late Modern period, see Mugglestone (1995: 107–150). The stigmatization of the alveolar pronunciation of <ing> was equally strong by the late 20th century, but here the story is more complex, involving social stratification. Walker was aware of the distinction between /n/ and /ŋ/ and provides evidence that the use of /ɪn/ for the *-ing* morpheme was condemned by some teachers, but he states that “our best speakers do not invariably pronounce the participle *ing*, so as to rhyme with *sing*, *king* and *wing*” (Walker, 1791: lxxxviii). The alveolar pronunciation was a marker both of lower-class and upper-class usage throughout the Late Modern period. By the early 20th century, the upper-class use of the alveolar was becoming a target of humor, but it can still be heard in the speech of very elderly, very conservative RP speakers.

Vowels

Such vowel changes as occurred in the Late Modern period largely involve the continuation of processes begun in the 16th and 17th centuries. In all cases, the earlier variants are still found in English regional accents, with many of the innovations still confined to RP and southern varieties. One of the clearest and most persistent markers of the ‘north–south divide’ in English accents, the presence or absence of /ʌ/ in *blood*, *cup*, *put*, etc. had already been established by the mid-18th century, as Walker points out:

If the short sound of the letter *u* in *trunk*, *sunk* etc., differ from the sound of this letter in the northern parts of England, where they sound it like the *u* in *bull* ... it necessarily follows that every word where this letter occurs must by these provincials be mispronounced’ (Walker, 1791: xiii).

What is new here is not the southern /ʌ/ so much as the attitude that the northerners’ lack of this phoneme marks them out as provincial. That other marker of the north–south divide, the pronunciation of the vowel in *bath*, *laugh*, *grass*, etc., has a more complex history. Evidence of lengthening of /æ/ in certain environments, mainly before voiceless fricatives, pre-consonantal (but not final) /r/, and /n/ followed by another consonant, as in *dance*, etc. occurs in the late 17th century, but, throughout the 18th and 19th centuries, the pronunciation with /ɑ:/ was not universally accepted. Walker tells us that although ‘Italian *a*’ had previously been heard in words such as *glass*, *fast* “this pronunciation seems to have been for some years advancing to the short sound of this letter, as heard in *hand*, *land*, *grand* etc. and pronouncing the *a* in *after*, *answer*, *basket*, *plant*, *mast*, etc. as long as in *half*, *calf* etc. borders very closely on vulgarity” (Walker, 1791: 10). This change seems to have begun as a lengthened [æ:] in the 17th century, and not to have become stigmatized until the lengthened vowel was retracted to [ɑ:]. The latter pronunciation is described as ‘drawling’ throughout the 19th century, and there is evidence of a pronunciation with [æ] or even [ɛ] by young ladies wanting to avoid the ‘vulgar’ [ɑ:]. Those who wished their pronunciation to be beyond reproach had to avoid both the ‘drawling’ [ɑ:] and the ‘mincing’ [æ] at least until the beginning of the 20th century, when Daniel Jones’s use of cardinal *ɑ* seems to have established this as the RP pronunciation. The lengthening of short *o* to /ɔ:/ in *off*, *cloth*, *cross*, etc., likewise began in the late 17th century and was considered ‘vulgar’ through most of the Late Modern period. Walker explicitly draws a parallel between the two vowels:

What was observed of the *a*, when followed by a liquid and a mute, may be observed of the *o* with equal justness. This letter, like *a*, has a tendency to lengthen when followed by a liquid and another consonant, or by *s*, *ss* or *s* and a mute. But this length of *o*, in this situation, seems every day growing more and more vulgar; and, as it would be gross, to a degree, to sound the *a* in *castle*, *mask* and *plant*, like the *a* in *palm*, *psalm*, &c. so it would be equally exceptionable to pronounce the *o* in *moss*, *dross* and *frost*, as if written *mawse*, *drawse*, and *frawst*. (Walker, 1791: xx)

As Mugglestone (1995: 231) points out, both [ɑ:] in *bath*, etc., and [ɔ:] in *off*, etc., were condemned as

‘vulgar’ throughout the 19th century, most probably because of their association with Cockney, and both became acceptable in 20th-century RP. However, while the former remains in present-day RP and southern accents of England, [ɔ:] in *off* is now a stereotype of very conservative, very upper-class RP, such as that of older members of the Royal Family.

The other vowel changes to be considered here could be regarded as the tail-end of the Great Vowel Shift. In words such as *face* and *goat*, the ME vowels had been raised to /e:/, /o:/, respectively, and these are the pronunciations described by Walker, who writes of the ‘long, slender’ sound of <a> and the ‘long, open’ sound of <o>, respectively. While the exact quality of the vowels described by Walker is open to debate (MacMahon, 1998: 450; Beal, 2004: 136–138), the pronunciations described are monophthongal. However, the first evidence for diphthongal pronunciations of both these vowels comes very soon after Walker’s first edition: MacMahon (1998: 459) points out that the first evidence for diphthongization of /o:/ comes from the Scottish orthoepist William Smith in 1795 and it is generally accepted that the first attestation of a diphthongal pronunciation of /e:/ comes from Batchelor (1809). In both cases, the diphthongal pronunciations are widely accepted in the 19th century, and are still found in RP and many other accents of present-day English.

Lexical Innovation

Given the external factors referred to above, we would expect the Late Modern period to be a time of lexical expansion, as all the new inventions and discoveries of the scientific age would require names, and English speakers encountered a wide variety of languages as a result of trade, exploration, and colonization. **Figure 1** is based on information provided by the *Chronological English dictionary* (Finkenstaedt *et al.*, 1970), which, in turn, took its data from the *Shorter Oxford English dictionary*. Most recent accounts of lexical innovation in English, such as Bailey (1996) and Görlach (1999, 2000), take their information from this source, even though it

has limitations. **Figure 1** shows a distinct downturn in lexical innovation in the mid-18th century, an increase in the 19th century, reaching a peak in the mid-19th century, and a tailing-off toward the 20th century. The apparent decline at the end of the 19th century can be dismissed as the information ultimately comes from the first edition of the *Oxford English dictionary*, which was being produced at the end of the 19th century and therefore did not tend to include innovations from this period. The *OED online* provides a more accurate picture: for instance, a search for the year 1890 under ‘first citations’ provides 1235 entries. The apparent trough in the 18th century can partly be explained by the relative neglect of 18th-century sources on the part of the original *OED* compilers, but it is also true that there was a tendency to resist innovation in this period. Authors such as Swift and Addison satirized the ‘affectation’ of slang words such as *mob* and *bamboozle* and the importation of French military terms such as *corps* and *c(h)arte blanche*. Objections to loanwords from French were voiced at various points in the 18th and early 19th centuries, when war between the two nations brought speakers of French and English into contact, and news of France’s superior military engineering into the English papers. Several reviews in 18th-century periodicals comment adversely on the influx of French military terminology, but even in times of peace, the excessive use of French words is condemned as an affectation. In 1771, the *Monthly Review* criticizes those who use *rôle* for *part* or *penchant* instead of *the passion of love*, stating that “the offended ear of the unfrenchified reader sickens at the sound” of these words. Of course, both these words are now accepted in English, but, like many French loans of the Later Modern period, they have not been fully anglicized. Apart from French, the other major sources of loanwords in the 18th century were Latin and, to a lesser extent, Greek. Publications such as Chambers *Cyclopaedia* brought information on new scientific classifications, discoveries, and inventions to a wide readership, introducing words coined from classical roots. Examples from 1753 are *aeronautics*, *azalea*, and *caldarium* from Latin, and

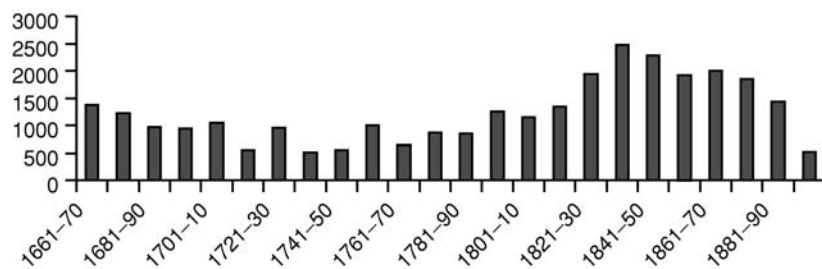


Figure 1 Numbers of first citations in the *Shorter Oxford English Dictionary* by decade 1661–1900.

aetiological, *eczema*, and *splenitis* from Greek: all of these were first cited in the 1753 supplement to Chambers *Cyclopedia*. The number and the proportion of new words formed from classical roots was to increase substantially in the 19th century: figures from *CED* for 1835 show more than two-thirds of the words first cited in that year having Latin or Greek etymologies (Beal, 2004: 25). Some 19th-century authors objected to what they saw as an excessive dependency on these sources. Richard Grant White complained:

In no way is our language more wronged than by a weak readiness with which many of those who, having neither a hearty love nor a ready mastery of it, or lacking both, fly readily to the Latin tongue or to the Greek for help in naming a new thought or thing or the partial concealment of an old one (1872, 22, cited in Bailey 1996: 141–142).

Despite such complaints, many of the scientific discoveries and inventions of the 19th century were given names coined from classical roots: examples from 1835 are *bifurcate*, *capilliform*, and *locomotory* from Latin and *creosote*, *phonograph*, and *silo* from Greek. The scientific discoveries of this century also led to a growth in the number of eponyms, as inventors or discoverers claimed a stake in posterity by having a new process or mineral named after them. The new science of geology provided many new words for minerals named after the geologist who first found them or the place they were found: examples first cited in 1835 include *bromlite*, *lanarkite*, *leadhillite*, *proustite*, *smithsonite*, *stromeyerite*, *troostite*, *uralite*, and *voltzite*. These are the major sources of lexical innovation in the 19th century, but what is also evident is that, as the Late Modern period progresses, words are imported into English from a wider variety of languages as exploration and colonization brought terms for the flora, fauna, and customs of the Americas, Asia, Africa and, last of all, Australasia. In 1835, for instance, the words *kiwi*, *rata*, and *tui* are first cited, all from Maori.

Bibliography

- Arnaud R (1983). 'On the progress of the progressive in the private correspondence of famous British people (1800–1880).' In Jacobson S (ed.) *1982 Papers from the second Scandinavian symposium on syntactic variation, Stockholm, 15–16 May*. Stockholm: Almqvist and Wiksell. 83–94.
- Austin F (1984). 'Double negatives and the eighteenth century.' *English Historical Linguistics* 3, 138–148.

- Bailey R W (1996). *Nineteenth-century English*. Ann Arbor: University of Michigan Press.
- Batchelor T (1809). *An orthoepical analysis of the English language*. London.
- Beal J C (1999). *English pronunciation in the eighteenth century: Thomas Spence's 'Grand repository of the English language.'* Oxford: Clarendon Press.
- Beal J C (2003). 'John Walker: prescriptivist or linguistic innovator?' In Dossena M & Jones C (eds.). 83–105.
- Beal J C (2004). *English in modern times*. London: Arnold.
- Cheshire J, Edwards V & Whittle P (1993). 'Non-standard English and dialect levelling.' In Milroy J & Milroy L (eds.) *Real English*. London: Longman. 53–96.
- Curme G O & Kurath H (1931). *A grammar of the English language, vol. III: syntax*. New York, Boston: D. C. Heath & Co.
- Denison D (1998). 'Syntax.' In Romaine S (ed.) *The Cambridge history of the English language, vol. IV: 1776–1997*. Cambridge: Cambridge University Press. 92–326.
- Dennis L (1940). 'The progressive tense: frequency of its use in English.' *Proceedings of the Modern Language Association of America* 55, 855–865.
- Dossena M & Jones C (eds.) (2003). *Insights into late modern English*. Bern: Peter Lang.
- Ellegård A (1953). *The auxiliary "do": the establishment and regulation of its use in English*. Stockholm: Almqvist & Wiksell.
- Finkenstaedt T, Leisi E & Wolff D (1970). *A chronological English dictionary*. Heidelberg: Winter.
- Görlach M (1999). *English in nineteenth-century England*. Cambridge: Cambridge University Press.
- Görlach M (2001). *Eighteenth-century English*. Heidelberg: Winter.
- Greenwood (1711). *An essay towards a practical English grammar*. London: R. Tookey.
- Holmberg B (1964). *On the concept of Standard English and the history of Modern English pronunciation*. Lund: Gleerup.
- Johnson S (1755). *A dictionary of the English language*. London: W. Strahan for J. & P. Knapton, T. & T. Longman et al.
- Jones C (1989). *A history of English phonology*. London: Longman.
- Lass R (ed.) (1999). *The Cambridge history of the English language, vol. III: 1476–1776*. Cambridge: Cambridge University Press.
- Leonard S A (1929). *The doctrine of correctness in English usage 1700–1800*. Wisconsin: Madison.
- Lowth R (1762). *A short introduction to grammar*. London: R. Dodsley.
- MacMahon M (1998). 'Phonology.' In Romaine S (ed.). 373–535.
- Marsh G P (1860). *Lectures on the English language*. New York: Burchfield.
- McKnight G H (1928). *Modern English in the making*. Reprinted in 1968 as *The evolution of the English language*. New York: Dover Publications.

- Mugglestone L (1995). *Talking proper*. Oxford: Clarendon Press.
- Palsgrave T (1530). *L'esclaircissement de la Langue Francoyse*. London: Richard Pynon.
- Porter R (2000). *Enlightenment: Britain and the creation of the modern world*. London: Penguin.
- Poutsma H (1928). *A grammar of Late Modern English* (2nd edn.). Groningen: Noordhoff.
- Raybould E (1998). 'How far was English syntax affected by Latin in the age of Dr Johnson?' In Ryden M et al. (eds.) *A reader in Early Modern English*. Frankfurt am Main: University of Bamberg studies in English Linguistics. 187–199.
- Romaine S (ed.) (1998). *The Cambridge history of the English language*, vol. IV: 1776–1997. Cambridge: Cambridge University Press.
- Sheridan T (1762). *A course of lectures on elocution*. London.
- Sheridan T (1780). *A general dictionary of the English language*. London: J. Dodsley, C. Dilly and J. Wilkie.
- Strang B M H (1970). *A history of English*. London: Methuen.
- Strang B M H (1982). 'Some aspects of the history of the be + ing construction.' In Anderson J (ed.) *Language form and linguistic variation: Papers dedicated to Angus Macintosh. Current issues in linguistic theory 15*. Amsterdam: Benjamins. 427–474.
- Sweet H (1873–4). 'The history of English sounds.' *Transactions of the Philological Society*, 461–623.
- Tieken Boon Van Ostade I (1982). 'Double negation and eighteenth-century grammars.' *Neophilologus* LXVI. 278–285.
- Tieken Boon Van Ostade I (1987). *The auxiliary 'do' in eighteenth-century English: a sociohistorical-linguistic approach*. Dordrecht: Foris.
- Tieken Boon Van Ostade I (2000). 'Of norms and networks: the language of Robert Lowth.' *Paper presented to the 11th International Conference for English Historical Linguistics, Santiago de Compostella, August 2000*.
- Visser Fr T (1963–1973). *An historical syntax of the English language* (3 parts). Leiden: Brill.
- Walker J (1791). *A critical pronouncing dictionary*. London: G. G. J. & J. Robinson and T. Cadell.
- Watts R & Trudgill P (2002). *Alternative histories of English*. London: Routledge.
- Webster N (1789). *Dissertations on the English language*. Boston: Isaiah Thomas & Co.
- Webster N (1828). *An American dictionary of the English language*. New York: S. Converse.
- Wright J (1898–1905). *The English dialect grammar*. Oxford: Henry Frowde.
- Wyld H C (1936). *A history of modern colloquial English* (3rd edn.). Oxford: Basil Blackwell (1st edn., 1920).

English, Middle English

J J Smith, University of Glasgow, Glasgow, UK

© 2006 Elsevier Ltd. All rights reserved.

The English of the period between the Norman Conquest of 1066 A.D. and the arrival of printing in England in 1476 is generally referred to as Middle English, as opposed to Old English (before 1066) and New or Modern English (after c. 1500). This terminology was established by the late-19th-century scholars Henry Sweet and Julius Zupitza. The dates are, of course, only useful signposts because the transitions between the three periods were gradual.

External History

William of Normandy's victory over the Anglo-Saxons was followed relatively swiftly by the imposition of Norman political and cultural hegemony throughout the kingdom. By William's death in 1087, the first two classes (estates) of medieval society, the clergy and nobility, were dominated by Normans.

This major change in England's social structure had a profound effect on the status of the English language, which had hitherto occupied a position unparalleled among the western European vernaculars. When the Normans arrived, they found a sophisticated society that had developed a distinctive vernacular culture. Spoken Old English consisted of a range of different varieties, strongly affected – especially in the north and east of England – by the Norse dialects of Viking settlers. In the written mode, however, one Old English dialect, the Late West Saxon in southwest England, had achieved the prestige associated with standard languages and was copied in various scriptoria, largely monastic, outside its area of origin. The Norman Conquest ended the prestige of Late West Saxon and, although texts continued to be copied in this standardized language for some time after 1066, dialectal variation and linguistic changes, hitherto not evidenced in written English, began to spread from the spoken to the written mode. The resulting language was Middle English, the earliest surviving example of which is probably the Final Continuation of the *Peterborough Chronicle* (see Clark, 1970).

Two languages replaced Late West Saxon in prestige: Latin and Norman French. The Conquest coincided with a revival of Latin learning in western Europe, and the Channel State of England and Normandy resulting from William's victory aided the transmission of this culture to Britain. Latin became the language of official record in England during the 12th and early 13th centuries, used for the *Domesday Book* (1086) and the Magna Carta (1215); it was also the literary language used by important 12th-century writers working in England, such as Geoffrey de Vinsauf, Alexander Neckham, and John of Salisbury.

Norman French, although the mother tongue of the invading elite, could not at first compete with Latin in all its functions. However, as Clanchy (1979: 168) pointed out, "contact with England, with its long tradition of non-Latin writing, may have helped to develop French as a written language"; and, from the 13th century onward, as Norman French developed in England into what modern scholars call Anglo-Norman, it began to be used for both official and literary purposes.

Throughout this period, English remained the primary language of the majority of the population of England, which peaked at just over 6 million in the middle of the 14th century. There is good evidence that the Norman aristocracy themselves had begun to speak English by the beginning of the 12th century, although French remained a necessary accomplishment for cultivated people; the change was encouraged by the loss of Normandy in 1204. English became increasingly widely used in the written mode as the Middle Ages progressed. However, its functions were parochial; it was used for local audiences and in the equivalent of primary education. The national functions of written language, until the very end of the period, were carried out by Latin and French; there was therefore no need of a national standard English. As a result, written English for much of the Middle English period manifested a high degree of variation of the kind now more generally associated with speech. There are, for instance, 143 distinct spellings for the item *such* recorded in the authoritative *Linguistic atlas of late mediaeval English* (McIntosh *et al.*, 1986), ranging from *schch* recorded in Norfolk to such forms as *swich*, *seche*, and *soche* to Kentish *zuyche* and Northern *swilk* and *slik*. It was only in the 15th century, with the rise of London English as a prestigious written language associated with the power and functions of the capital city, that a new standard written English emerged. Even then, extensive written variation remained into the 16th century, at which point the early printers began to provide authoritative norms for private use.

Internal History

Increased vernacular literacy from the 13th century onward means that, compared with Old English, a great deal of Middle English writing in contemporary manuscripts has survived. This material forms the primary evidence for Middle English.

Graphology

Given the variety of written Middle English, it is unsurprising that handwriting styles differed diachronically and diatopically and even in the work of single scribes. In general, the Old English insular script was replaced by continental styles; usages hitherto restricted to Latin texts began to be adopted when writing English. Cursive handwriting developed as a practical response to increasing literacy in both Latin and the vernaculars.

The Middle English alphabet was almost identical with that of present-day English. The Old English letters <æ> (ash), <ð> (edh), and <ƿ> (wynn) disappeared early in the Middle English period, being replaced by <a, e>, <th, þ>, and <w>, respectively. Old English runic <þ> (thorn) was retained alongside its ultimate replacement, <th>, for some time, although commonly realized as <y>, especially in northern varieties; when printing arrived, <þ> largely disappeared, but was retained (written <y>) in a few contexts where ambiguity did not arise, for example, <ye> 'the.' A modified form of the Old English insular g, <ȝ> (yogh), was retained by many scribes, commonly to represent [x, j]; the French habit of using <ȝ> to realize [z] was also adopted by many Middle English scribes, in addition to <z>, which was adopted from Latin usage. <g>, used in Anglo-Saxon times for copying Latin, was adopted to represent [g]. <c, k, q> developed their present-day usage during this period as a result of the adoption into English of practices used for writing French. <h> was used as a diacritic to indicate a modification of the letter it followed: hence, the development of <sh> (earlier <sch>) for Old English <sc>, <gh> for <ȝ> and <wh> for Old English <hw>. The letters <u, v> were used interchangeably to represent both vowel and consonant, with <v> generally being used initially and <u> elsewhere.

In Old English, <y> had represented [y], but that vowel was unrounded in many late Old English dialects, merging with Old English <i> [i] (In western dialects, the rounding was retained, but the vowel seems to have retracted to merge with <u> [u].) <y> then came to be used interchangeably with <i>, especially in environments where contemporary handwriting could be confusing (e.g., before or

after <m, n, u>). <o> was used for <u> in similar circumstances.

The combinations <ou, ow> were adopted from French usage in place of <u> in words such as *how* and *brown* (cf. Old English *hu, brun*). Toward the end of the period, southern varieties frequently indicated vowel length by doubling the letters representing them (e.g., *good, feed*); however, in northern and Scottish varieties, long vowels were often indicated by the addition of <i> (e.g., *guid* 'good'). As inflexional-*e* fell out of use at the end of the period (see the section on Noun Phrases), it too became available as a diacritic mark to indicate a long vowel, as in *life* (cf. Old English *lif*).

Phonology

Knowledge of Middle English phonology derives from the analysis of rhyming verse, reconstruction from later and earlier states of the language, and the interpretation of orthography. There is no thorough contemporary account of Middle English pronunciation comparable to the Old Icelandic *First grammatical treatise*, although there were scribes whose orthographic practices demonstrate considerable sophistication in handling the complexities of sound-symbol mapping (e.g., Orm, the author and scribe of *The ormulum*, c. 1200). The inventory of phonemes in Middle English varied diatopically and diachronically, so the following account is very general.

Consonants The phonemic consonants in Middle English were /p, b, t, d, k, g, tʃ, dʒ, f, v, θ, ð, s, z, ʃ, h, m, n, l, r, w, j/. The major structural difference between Old and Middle English was in the fricatives; voiced and voiceless fricatives were allophones in Old English, but were phonemicized in the Middle English period.

Vowels The Old English distinction between long and short stressed vowels remained important, but ceased to be phonemic as the Middle English period progressed; there is indirect evidence that the short vowels tended to have more open pronunciations than their long equivalents from the beginning of the 13th century onward. By the beginning of the 15th century, London English seems to have had the following inventory: /i, ɪ, e, ɛ, ɛ:, a:, ʌ, ɔ:, o, u, ʊ/; /i, e, o, u/ were long vowels where length had ceased to be phonemic. Quantitative changes, some already under way in the late Old English period, meant that short vowels tended to appear in closed syllables, whereas long vowels tended to appear in stressed open syllables; thus, contrastive distribution was lost.

The major difference between Old and Middle English vocalism was in diphthongs; the Old English diphthongs monophthongized, and new diphthongs arose from vocalizations of Old English [w, j, h]. French loanwords supplied the inventory with two new diphthongs, /ʊɪ, ɔɪ/.

In unstressed vowels, the Old English qualitative distinctions were already becoming obscured by the late Anglo-Saxon period. This process continued in Middle English; /ə/ was the most common vowel, although /ɪ/ spread from the north, and /ʊ/ (indicated by spellings such as <-us, -ud>) seems to have been characteristic of the west.

Grammar

Inflexion In Middle English, inflexions are not as functionally important as they were in Old English. Many roles played by inflexions in Old English were taken over by prepositions and a more fixed word order (see section on Word Order). There were, of course, dialectal differences. In general, innovations in morphology spread from the north of England to the south, so features found in late Northumbrian Old English (e.g., the 10th-century *Lindisfarne gospels gloss*) first appear in the south in Middle English texts.

Noun Phrase The masculine/feminine/neuter grammatical gender system of Old English disappeared in the Middle English period. Although inflexional distinctions remained in the personal pronouns of Middle English, these were assigned according to natural gender. This pattern was already becoming established in late Old English, when *wif* 'woman,' a neuter noun, was occasionally referred to by the pronoun *heo* 'she.' The case system had already been subject to syncretism in pre-Old English (as in all Germanic languages); it was during the Middle English period that it largely disappeared. Only inflexional markers of plurality and possession remained. Modifiers such as determiners and adjectives ceased to be marked for agreement with their head words.

In place of the four major and four minor noun declensions of Old English, Middle English had in general the modern pattern, that is, the simple addition of -(e)s to mark plurality or possession, although there were rather more relicts of the *micel/children*-type than there are in present-day English. The forms in -s derive from the Old English strong masculine paradigm.

Traces of the old adjectival strong/weak (indefinite/definite) distinction lasted until the late 14th century in southern England, indicated by the use of -e in some

paradigms; thus the poet Geoffrey Chaucer (d. 1400) distinguished *the olde man* and *the man is old*. However, the system had died out everywhere by the beginning of the 15th century, and *-e* was subsequently used as a diacritic of length (see section on Graphology).

The Old English system of inflexionally differentiated determiners collapsed, and the present-day system (*the/this/that/these/those*) gradually emerged. An indefinite article *a(n)* appeared, derived from Old English *an* 'one.' Perhaps the most radical changes took place in the pronominals; the Old English dual pronouns disappeared, and a number of other forms (e.g., Old English *hie* 'they,' *heo* 'she'), which had ceased to be distinctive because of sound changes, were replaced by variants that were available in the lexicon through contact with Old Norse.

Verb Phrase The strong/weak/irregular paradigms of Old English remained, with numerous analogical reassignments. Verb inflexions, although reduced and reorganized, were retained to mark agreement between subject and predicator.

Complex verb phrases arose in place of some Old English simple verb phrases. Old English distinguished between *bundon* 'bound' (preterite plural indicative) and *bunden* 'may have bound' (preterite plural subjunctive); the obscuration of the inflexional distinction led to the replacement of the formal subjunctive by complex verb phrases with *may* and *might* (Old English *mæg, mihte*, etc., 'can, could'). *Shall* and *will* were increasingly used as auxiliary verbs indicating future time rather than as lexical verbs signaling obligation and volition, respectively. Toward the end of the Middle English period, a complex verb phrase with *do* emerged in affirmative declarative constructions (e.g., *she did eat, it doth illuminate*); this construction has not survived into present-day use.

Other characteristic Middle English innovations are the development of phrasal verbs, such as *put up* and *stand by*, and an extension of the use of impersonal verbs, such as *me thinketh* 'it seems to me.' The latter usage has disappeared from present-day use, but the former construction is still common, especially in informal styles.

Word Order In word order, there is a noticeable extension in Middle English of the subject-verb-object sequence from affirmative main clauses not beginning with an adverbial to main clauses beginning with an adverbial and to subordinate clauses, where the Old English orderings were prototypically adverbial-verb-subject-object and subject-object-verb, respectively (as in present-day German).

Lexicon

There are three main sources of loan words into English during this period: Norse, Latin, and French. (For useful lists, see Serjeantson, 1935.)

Norse Loans Many Norse words were actually taken into English in the late Anglo-Saxon period, but in general they are hidden by the standardized written mode. Most Norse loans express very common concepts (e.g., *bag, bull, egg, root, ugly, wing*), and it is notable that Norse has supplied such basic features as the third-person plural pronouns *they, their, and them* (cf. Old English *hie, hiera, him*).

Latin Loans A number of Latin words came directly into English during the Middle English period, largely as learned words carried over in the translation of Latin texts (e.g., *omnipotent* and *testament*). However, the great wave of Latin borrowings into English took place from the 15th century onward, with the rise of humanism, and this is therefore a feature of Early Modern English.

French Loans Up to the 13th century, recorded borrowings from French into English are few and were generally restricted to the registers of government (e.g., *justice, obedience, mastery, prison, service*). From the beginning of the 13th century to the end of the Middle English period, however, French words entered the English lexicon in large numbers. As contact with Normandy was lost, Central French, not Anglo-Norman, became the main source of these words. The range of domains covered by these words is vast (see Serjeantson, 1935: 12–156). This surge was socially driven; although the higher social classes did not speak French as their mother tongue by this time, French retained its social cachet, and the use of French expressions was an obvious way of signaling a higher social position. Even in present-day English, French-derived vocabulary is often stylistically marked as of a higher register; compare the difference in meaning between high-style *commence* and neutral *begin*.

French also affected word formation. On the one hand, compound forms, characteristic of Old English, were frequently replaced by simple borrowings (cf., *brekan* 'break,' *forbrekan* 'destroy'). On the other hand, French suffixes and prefixes were applied to native stems (e.g., *knowable, unspeakable*).

Typical Middle English

The heterogeneity of written Middle English means that a typical specimen of Middle English is not to be

had. The following text from Scragg (1974: 31–32) is a version of the Lord's Prayer in a Central Midlands dialect of the 1380s (MS London, British Library, Royal 1.B.vi); it exemplifies some of the features previously discussed.

Oure fadir, þat art in heuenys, halewid be þi name. þi kyngdom come. Be þi wille don as in heuene and in erþe. ȝiue to vs yis day oure breed ouer oþer substaunse. And forȝiue to vs oure dettes, as and we forȝiuen to oure dettouris. And leede vs not into temptacioun, but delyuere vs from yuel.

Modern Work on Middle English

The two most important recent single publications in Middle English studies are the *Middle English dictionary* (Kurath *et al.*, 1951–2001; also online), and *A linguistic atlas of late Mediaeval English* (LALME; McIntosh *et al.*, 1986). The former allows for a much more detailed investigation of the Middle English lexicon than has been possible hitherto. The parallel completion of the *Dictionary of the older Scottish tongue* (Craigie *et al.*, 1931–2002) means that the resources for the study of medieval lexicons of Britain are now massively enhanced. In combination with the continued evolution of the *Oxford English dictionary*, and the imminent completion of the *Historical thesaurus of English*, these publications will allow a great leap forward in the diachronic study of vocabulary and its structure.

LALME has opened up a mass of unpublished evidence for investigation, and also, by localizing so many texts to particular places, means that a much greater range of dialectal grammars can be constructed than has yet been achieved. Follow-up projects on Early Middle English and Older Scots are currently under way.

The indispensable foundation, however, for the study of Middle English remains: the editing of texts. Some fashions in literary (as opposed to linguistic) study, however, militate against the usefulness of this enterprise; the modern practice: even in scholarly editions: is to make numerous silent decisions in editing Middle English texts. Such decisions disguise important linguistic features such as punctuation, marks of abbreviation, and even spelling.

Alongside these developments, theoretical work continues. Probably the most important descriptive milestones are the major histories, such as the second volume of the *Cambridge history of the English language* (Blake, 1992) and the single-volume *Oxford history of the English language* (Mugglestone, 2005). Explanatory work is necessarily more controversial; probably the single most influential work on the

historical study of English, with a special focus on Middle English, remains Samuels (1972).

Bibliography

- Barber C L (1993). *The English language: a historical introduction*. Cambridge, UK: Cambridge University Press.
- Blake N (ed.) (1992). *The Cambridge history of the English language 2: 1066–1476*. Cambridge, UK: Cambridge University Press.
- Baugh A C & Cable T (2002). *A history of the English language* (5th edn.). London: Routledge.
- Brunner K (1970). *An outline of Middle English grammar*. Johnston G (trans.). Oxford: Blackwell.
- Clanchy M T (1979). *From memory to written record: England 1066–1307* (2nd edn., 1993). London: Arnold.
- Clark C (ed.) (1970). *The Peterborough chronicle*. Oxford: Clarendon Press.
- Craigie W *et al.* (eds.) (1931–2002). *A dictionary of the older Scottish tongue*. Chicago, IL: Chicago University Press/Aberdeen, UK: Aberdeen University Press/Oxford: Oxford University Press.
- Denison D (1993). *English historical syntax*. London: Longman.
- Fisiak J (1968). *A short grammar of Middle English*. Warsaw: Panstwowe Wydawn., Naukowe.
- Horobin S & Smith J J (2002). *An introduction to Middle English*. Edinburgh, UK: Edinburgh University Press.
- Jones C (1972). *An introduction to Middle English*. New York: Holt, Rinehart and Winston.
- Jones C (1987). *Grammatical gender in English: 950 to 1250*. Beckenham, UK: Croom Helm.
- Jordan R (1974). *Handbook of Middle English grammar: phonology*. Crook E J (trans.). The Hague: Mouton.
- Kurath H *et al.* (eds.) (1951–2001). *Middle English dictionary*. Ann Arbor, MI: University of Michigan Press.
- Lass R (1987). *The shape of English*. London: Dent.
- McIntosh A, Samuels M L & Benskin M (1986). *A linguistic atlas of late mediaeval English*. Aberdeen, UK: Aberdeen University Press.
- McIntosh A, Samuels M L & Laing M (1988). *Middle English dialectology: essays on some principles and problems*. Aberdeen, UK: Aberdeen University Press.
- Machan T W (2003). *English in the middle ages*. Oxford: Oxford University Press.
- Millward C (1988). *A biography of the English language*. New York: Holt, Rinehart and Winston.
- Mugglestone L (2005). *The Oxford history of the English language*. Oxford: Oxford University Press.
- Mustanoja T F (1960). *A Middle English syntax I: Parts of speech*. Helsinki: Societe Neophilologique.
- Parkes M B (1979). *English cursive book hands 1250–1500* (2nd edn.). London: Scolar Press.
- Samuels M L (1972). *Linguistic evolution, with special reference to English*. Cambridge, UK: Cambridge University Press.
- Scragg D (1974). *A history of English spelling*. Manchester, UK: Manchester University Press.
- Serjeantson M S (1935). *A history of foreign words in English*. London: Routledge.

- Smith J J (1996). *An historical study of English*. London: Routledge.
- Smith J J (1999). *Essentials of early English*. London: Routledge.
- Strang B M H (1970). *A history of English*. London: Methuen.

- Warner A (1982). *Complementation in Middle English and the methodology of historical syntax*. London: Croom Helm.
- Wright J & Wright E M (1928). *An elementary Middle English grammar*. London: Oxford University Press.

English, Old English

C L Allen, Australian National University, Canberra, Australia

© 2006 Elsevier Ltd. All rights reserved.

Old English is the development of the language which was introduced into Britain by Germanic invaders in the 5th century A.D. and became the dominant language of what is now England excluding Cornwall, and also southern Scotland. Old English evolved gradually into Modern English, and any date given to delineate the different stages of the language must be somewhat arbitrary. The period covered by Old English spans c. 425 to c. 1100, and the language underwent extensive changes in this period. The comments made below about the phonology and grammar of Old English must be understood to apply to some of this period, but not necessarily all. By 1100, the accumulation of changes had become so great that it is appropriate to speak of Middle English.

While we do not know what the earliest Germanic people coming to Britain called their language, we find it referred to as *englisc* c. 1000, and *Engla land* (land of the Angles) used for the area where it was spoken. It is impossible to estimate the number of speakers throughout the Old English period, but William the Conqueror's Domesday Survey (1086) suggests an Anglo-Saxon population of between one and two million.

It is customary to recognize four basic dialects of Old English, distinguished from each other most prominently by phonological features but also by morphological and lexical peculiarities: Northumbrian (northern), Mercian (midlands), Kentish, and West Saxon (southwestern). It is tempting to attribute the main dialect differences to the premigration period on the continent, following the tradition beginning with the famous description of the Germanic invaders in Bede's *Ecclesiastical history of the English people*. Bede, writing in Latin in 731, says that the invasion force consisted of three tribes – Angles, Saxons, and Jutes – who founded separate kingdoms. But the majority opinion of modern scholars stresses

the complexity of political and social factors that presumably contributed to the creation of regional dialects in addition to the variation that would have come with the earliest Germanic settlers.

Our knowledge of Old English dialects and how they developed is limited by the sparseness of records in English until the 9th century. Because this was a period of political hegemony by Wessex accompanied by a burst of literary activity associated with Alfred the Great (who reigned from 871 to 899), the majority of surviving Old English manuscripts were written in the West Saxon dialect, and West Saxon became a sort of standard for writing in most areas of the country. For this reason, West Saxon is what is normally taught as Old English. Standard Modern English is not in a straight line of descent from West Saxon, but derives mainly from southeastern Mercian varieties for which our information in the Old English period is scanty.

Genetic Relationships

Comparison of Old English with other languages shows that it belonged to the Germanic subgroup of the large Indo-European family, which includes Old Saxon, Old High German, Old Frisian, Gothic, and Old Norse. An example of a major phonological feature that sets the Germanic languages off from the other Indo-European languages is the Germanic shift of stress to the initial syllable of every word (excluding certain prefixes). Using the comparative method, we can reconstruct a good deal of the vocabulary and morphology of Germanic; for example, Old English, Old Saxon, and Old Frisian share the form *ūs* 'us', which can be shown to have descended from the same form **uns* in Germanic. The word shows up as *uns* in Gothic and Old High German and *oss* in Old Norse.

Such comparison indicates that Old English, Old Frisian, and Old Saxon shared some sound changes which differentiated them from the other Germanic languages. In the example just cited, the nasal *n* has been lost before the fricative *s*, with compensatory lengthening of the vowel, in these languages, and similarities of this sort have caused Germanic

scholars to refer to these three languages as 'Ingvaeonic' or 'North Sea' Germanic. A smaller grouping of Anglo-Frisian can be made within the Ingvaeonic group. Old English and Old Frisian were the only Germanic languages that fronted the Indo-European back vowel *a* to front *æ*. The Ingvaeonic languages share with Old High German some consistent similarities that differentiate them from the Scandinavian languages such as Old Norse on the one hand (where assimilation of the nasal to the fricative gives *oss* from **uns*) and Gothic on the other, and so a tripartite division of the Germanic languages into West, North, and East is traditional. This picture is complicated by the fact that the speakers of the West and North Germanic languages maintained a reasonable degree of contact with each other. Even while their languages were undergoing separate development, they retained a certain amount of unity and shared innovations. The result is that Old English was more like Old Norse than it was like the East Germanic Gothic.

Written Records

We can never know for certain how Old English sounded or everything about the grammar of the language, but there is a lot we do know. We cannot know the exact quality of the vowel in what was written as *mus* ('mouse'), for example, but we can tell that it must have been a long, high, back vowel. For one thing, comparison of the sister Germanic languages and the further history of the word in English (which saw the vowel become a diphthong) makes this the only candidate. For another, the Old English scribes adapted the Latin alphabet to represent their language, and it would have been perverse to use a letter to represent a sound that was entirely divorced from the sound it was used to represent in Latin.

Old English written records do not start appearing until c. 700 in quantities useful for studying the language of the time and its previous development from common Germanic. Prior to this, there are some inscriptions in the runes that the Germanic tribes used before their conversion to Christianity, when the Latin alphabet was adopted. The earliest of these dates to the 4th or early 5th century. However, runes were mostly used for inscriptions, not literary activity. The runic inscriptions found in Britain are also few in number, although they provide us with very important information (sometimes open to more than one interpretation). We also have a copy of the law code of Æthelberht of Kent (d. 616), but it is contained in a manuscript from a much later period. The fact that anyone thought it worthwhile to write these laws down in English sets Old English apart from most

vernacular languages of such an early period, when Latin was the only language used for such subject matter in much of Europe.

The earliest writings in the Latin-based alphabet, adapted slightly to deal with the sounds of Old English, are mainly glosses of Latin. These are invaluable sources of information about the phonology and morphology of the earliest period of Old English, but they cannot teach us much about Old English syntax. It is only in the time of Alfred the Great that materials become abundant. We have more remains of Old English than any other Germanic language of such an early period, covering poetry, religious material, medicine books, a grammar of Latin for English speakers, the *Anglo-Saxon chronicle*, and more, amounting to approximately three million words. Scholars are well served with electronic corpora, of which the Helsinki Corpus (see Kahlas-Tarkka *et al.*, 1993) was the first. Such corpora have become an invaluable tool in the study of Old English morphology and syntax in particular.

From Germanic to Old English: Phonology

One of the most striking features of Old English phonology is that assimilation is found in more than one guise. Assimilation is found in most Germanic languages to some extent, but Old English underwent its own special assimilatory processes. A process usually known as *i*-umlaut is one of the most important developments of the prehistoric Old English period when Old English had separated from its sister languages but was not yet written down. The essence of this process was that a high front vowel or palatal glide in one syllable caused a back vowel in the preceding syllable to become a front vowel, to oversimplify greatly. This is a type of anticipatory assimilation, in which the speaker anticipates the vowel of an upcoming syllable by moving the tongue into a front position too early. An assimilation of an even earlier period involved the palatalization of velar consonants that were adjacent to front vowels or a palatal glide in the same syllable. Palatal stops have a tendency to turn into affricates, and this is what happened in Old English, with the result that Germanic **dik* has come down to us as *ditch*.

The importance of syllable structure in Old English is seen in a process that deleted a final vowel after 'heavy' syllables, which either contained a long vowel or ended in a consonant cluster. Thus, the nominative plural form was *scipu* for 'ship' because the base *scip* had a short vowel and ended in a single consonant, while *word* was either nominative singular 'word' or plural 'words' because the consonant cluster *rd* caused deletion of the plural suffix *-u*.

One interesting difference between Old English and the language that descended from it is that voicing was not a distinctive feature for most fricatives. In Modern English, [f] and [v], written as *f* and *v*, are perceived as two distinct sounds, as the difference between *fat* and *vat* illustrates. But in Old English, they were perceived as variants of essentially the same sound or phoneme. The labiodental fricative was pronounced as voiced [v] when surrounded by voiced sounds (such as vowels), but pronounced as voiceless [f] otherwise; e.g., although *deofol* was written with an *f*, it was pronounced more like modern *devil* than the spelling suggests.

Effects of Language Contact

Some everyday words borrowed from Latin such as *cēse* ‘cheese’ (from Latin *cāseus*) bear evidence of early sound changes and are likely to have entered the language prior to the invasion of Britain. The conversion to Christianity brought writing in the Roman alphabet to the Anglo-Saxons (with the help of Irish missionaries). However, Anglo-Saxon churchmen preferred to use native resources for translating ecclesiastical concepts, e.g., *halig gāst* ‘holy ghost’ as the translation of *spiritus sanctus*; *spirit* was not used until the Middle English period.

Since Britain was inhabited by Celts before the Germanic invasion, one might suppose that Old English would show a great deal of Celtic influence, but the number of Celtic words borrowed into Old English, other than place names such as *Temes* ‘Thames,’ is exceedingly small. This is a pattern typical of an invasion, c.f. the retention of aboriginal place names by English settlers in the United States and Australia. Expert opinion varies greatly as to the number of Germanic migrants and the displacement of the Celts, but whether it was a case of language shift by the Celtic population or the spread of the language by the spread of the Germanic population, there is no convincing evidence of substratum Celtic influence on Old English grammar or phonology.

In contrast, there can be no question of very intimate contact with Scandinavian speakers, particularly in the northeast of the country, due to the Scandinavian raids that began in the 8th century, which later became determined attempts at settlement and culminated in Danish rule of the entire country by Cnut (1017–1035). Such contact cannot fail to have linguistic consequences, although the nature of those consequences is a matter of debate. Our uncertainty is exacerbated by the paucity of written records from the areas where contact with Scandinavian speakers was the greatest in Old English. Many of the consequences in English do not appear in texts until the very late Old English or

Middle English period. However, the inflectional system of the *Lindisfarne Gospels*, an interlinear gloss that was added to a Latin text of the gospels in the 10th century in an area that had been under Scandinavian domination for more than a century, exhibits considerable modification of the inflectional system presented below. Many of these changes, such as the generalization of *-es* as the genitive singular inflection to declensional classes where it historically does not belong, are harbingers of more general changes of the Middle English period. This suggests that contact with Scandinavian speakers played an important role in shaping the dialects where contact was the greatest, with ramifications for the later development of Standard English. However, it is likely that in many instances, the effect of contact was essentially to take the brakes off linguistic changes that were already in progress.

With such closely related languages, it can be difficult to tell whether a word that is only recorded in the late Old English period or later is of native origin or a Scandinavian borrowing, but it is certain that large numbers of everyday words such as *take* which replaced native *niman* are of Scandinavian origin.

French influence in English mostly dates from after the Norman Conquest of 1066 and so mostly lies outside the Old English period, but contact with French did not begin with the Conquest; especially notable is that there was a strong Norman presence at the court of Edward the Confessor (reigned 1042–1066). Several French loan words entered the language in the Old English period, including *prūd* ‘proud’, first attested c. 1000.

Grammatical Characteristics

Modern English still exhibits much of the lexicon and many of the Germanic structural characteristics of Old English, but Old English is so different from Modern English that it must be learned as a foreign language. One of the primary distinctions between Old English and later English lies in the more elaborated inflectional morphology of Old English. For example, nouns and their modifiers were inflected according to two numbers (singular and plural), three genders (feminine, masculine, and neuter) and four cases (nominative, accusative, genitive, and dative). The distinctions were mainly made by suffixation, but also by mutation of the stem vowel, as in feminine nominative/accusative singular *bōc* ‘book’ but dative singular *bēc*. An instrumental case was distinguished to a limited extent, but it had mostly merged with the dative case, and a prepositional phrase was more commonly employed than the instrumental case. Here, as in many other places,

prepositional phrases were already being used to supplement case marking. Few forms were unambiguously markers of a single combination of grammatical features; most masculine nouns did not distinguish the nominative from the accusative in either the singular or the plural, while most feminine nouns made this distinction in the singular but made no distinction between the non-nominative cases. Modifiers of the nouns, especially determiners, distinguished grammatical features more consistently, e.g., the masculine 'the stone' is *se stān* (nominative), *þone stān* (accusative), *þæs stānes* (genitive) and *þām stāne* (dative). Fewer distinctions were made in the plural even with the demonstratives, e.g., *þā stānas* 'the stones' could be nominative or accusative.

The more elaborate case marking system of Old English made a freer order of constituents possible than is found in Modern or even Middle English. 'The man killed the king' could be expressed as either *se man acwealde þone cyning* or *þone cyning acwealde se man*, because *se man* was nominative, and identified the subject wherever it was positioned, and the accusative form *þone cyning* signaled the object. Already in Old English, however, the order Subject-Object (as in the first of our examples) was much more frequent than Object-Subject. This order was the usual one even when case marking would have made the grammatical relations clear without the help of word order.

The position of the verb was also variable. Old English exhibits the preference for verb-second order in main clauses and verb-final order in subordinate clauses found in the other early Germanic dialects, but in Old English, verb-second position in main clauses was not as regular as it is in, for example, Modern Dutch, and only about 50% of subordinate clauses have verb-final order. Order within the noun phrase was also variable but far from free, with the prenominal order Determiner Adjective Noun familiar from

Modern English already the unmarked pattern. The postnominal order found in *men þa leofestan* 'most beloved people' (lit. 'people the belovedest') was mostly limited to vocatives and poetry, and is presumably a relic of more widespread postnominal positioning of modifiers in Germanic. The scene was set for further dependence on word order in the Middle English period.

Bibliography

- Bede (1969). *Bede's ecclesiastical history of the English people*. Colgrave B & Mynors R A B (eds., trs.). Oxford: Clarendon Press.
- Campbell A (1959). *Old English grammar*. Oxford: Clarendon Press.
- Hogg R M (1992). *The Cambridge history of the English language vol. I: The beginnings to 1066*. Cambridge: Cambridge University Press.
- Hogg R M (2002). *An introduction to Old English*. Oxford: Oxford University Press.
- Kahlas-Tarkka L, Kilpiö M & Österman A (1993). 'Old English.' In Rissanen M, Kytö M & Palander-Collin M (eds.) *Early English in the computer age: explorations through the Helsinki Corpus*. Berlin: Mouton de Gruyter. 21–32.
- Lass R (1994). *Old English: a historical linguistic companion*. Cambridge: Cambridge University Press.
- Mitchell B (1985). *Old English syntax*. Oxford: Clarendon Press.
- Mitchell B & Robinson F C (2001). *A guide to Old English*. Oxford, Malden, MA: Blackwell Publishers.
- Quirk R, Wrenn C L & Desks S E (1994). *An Old English grammar*. DeKalb: Northern Illinois University Press.
- Robinson O W (1992). *Old English and its closest relatives: a survey of the earliest Germanic languages*. Stanford: Stanford University Press.
- Townend M (2002). *Language and history in Viking age England: linguistic relations between speakers of Old Norse and Old English*. Turnhout, Belgium: Brepols.

English, Variation in Nonnative Varieties

Z Bao, National University of Singapore, Singapore

© 2006 Elsevier Ltd. All rights reserved.

Native vs. Nonnative Varieties

Native and nonnative varieties of English are distinguished on the basis of the sociolinguistic environment in which they take root. Native varieties are found in North America, Australia, and New Zealand, places

that saw large-scale settlement by English-speaking people. Nonnative varieties emerge in former British or American colonies in South and Southeast Asia and parts of Africa, where there has never been a sizable English-speaking settlement, and English is spoken along with the languages of the local populations. From the perspective of genetic linguistics, native varieties are the product of normal parent-to-child transmission in that both the grammar and the vocabulary are transmitted from the same parent language

(Thomason and Kaufman, 1988). There is little structural difference among them. Nonnative varieties are more complicated. Though the vocabulary is largely English, the grammar exhibits significant restructuring under the influence of indigenous languages. Given their unique sociolinguistic histories, nonnative varieties are not typologically homogenous. Because of the presence of linguistic features appropriated from indigenous languages, they are often referred to as ‘indigenized,’ ‘nativized,’ or ‘contact’ varieties, or as ‘New Englishes.’

Nonnative varieties are distinguished from English-lexified pidgins and creoles, also on sociolinguistic grounds. However, the internal variation within a nonnative variety is analogous to the post-creole continuum, ranging from basilect to acrolect. Unlike the basilectal subvarieties, the acrolect does not exhibit the effect of grammatical restructuring and serves as the local standard. It is in effect a native variety, and ‘nonnative’ applies to the basilectal subvarieties.

Native vs. Nonnative Speakers

‘Native’ is also used to describe the order of language acquisition: a ‘native’ language is the first language acquired by a ‘native’ speaker. The acquisitional status of nonnative varieties of English deserves comment. Conditioned by different postcolonial experiences, they followed separate developmental paths (Schneider, 2003). Malaysia and Singapore offer an interesting case study. In Malaysia, Malay is the national language, and English remains the language of the elite. In Singapore, the government adopts an English-centered language policy. English is the working language of government and, more importantly, the medium of instruction in schools. An English diglossia has emerged, with the nonnative variety – Singapore English – as L, and the superposed, acrolectal variety as H. Increasingly Singapore English is acquired as a first, if not the first, language (Kwan-Terry, 1991; Gupta, 1994). Given the right sociolinguistic conditions, a nonnative variety can acquire native speakers and become the mother tongue.

Variation

Variation in nonnative varieties of English is usually measured against the grammatical norm of the native variety. The focus is placed on linguistic neologisms and their possible origins. Variation can also be approached in terms of the usage patterns of linguistic variables as conditioned by context of use. But this line of enquiry is woefully lacking in the literature and is usually subsumed in the more common studies of lectal variation, conditioned by speaker proficiency.

Linguistic neologisms can be found in all levels of language.

Phonetics and Phonology

Two noticeable innovations among the English consonants have to do with the dental fricatives (*thin*, *this*) and the aspirated voiceless stops (*pot*, *top*, *cop*). The dental fricatives are replaced by *t* and *d*, respectively, and aspiration of the voiceless stops is lost. These innovations illustrate two basic mechanisms of sound interference: direct substitution and change in phonological contrast. In Singapore English, there has been a further development in the treatment of the dental fricatives: they are pronounced as *t/d* in syllable-initial position, but as *f* in syllable-final position (*healthy* [-t-] vs. *health* [-f]).

The change in the vowel system is more drastic. The typical vowel inventory of a nonnative variety consists of five or seven vowels. The additional two vowels in the larger inventory are traceable to the diphthongs in *bait* and *boat*. A plausible explanation is that the five-vowel inventory is due to simplification in phonological contrast. Table 1 displays the result.

The five-vowel inventory emerges when length is no longer phonemic and the high–mid–low contrast is reduced to high–low. The loss of phonological contrast may be caused by contact with indigenous languages or by internal drift.

Lexicon and Morphology

The lexicon is a depository of words and is the part of language that is the most susceptible to external influence. Lexical borrowing is commonplace. Not surprisingly, nonnative varieties borrow words from the languages in their contact environment: *dhobi* ‘washer man’ (from Hindi) in India, *kampong* ‘village’ (from Malay), and *kaypoh* ‘nosy’ (from Chinese [Southern Min Chinese]) in Malaysia and Singapore. It is not easy to differentiate this sort of direct borrowing from code-mixing or code-switching, which are common phenomena in multilingual communities.

New meanings may develop. In Southeast Asia, to *gostan* (< *go stern*) is to change direction; an *alphabet* is a letter (*English has 26 alphabets*), and a *parking lot* is a space in a *car park*. A more subtle change

Table 1 Simplification and vowel inventory

| RP | | Nonnative | | Examples | |
|------|------|-----------|---|-----------------|-------------------|
| i:/ɪ | u:/ʊ | i | u | <i>beat/bit</i> | <i>boot/put</i> |
| ɛ | ɔ:/ɒ | e | o | <i>bet</i> | <i>caught/cot</i> |
| æ | ɑ:/ʌ | | a | <i>bat</i> | <i>cart/cut</i> |

involves the lexical semantics of words. Take *win* and *admit*. In Euro 2004, Greece played Portugal and won. To report this in Singapore English, you can say *Greece won*, *Greece won the game*, or *Greece won Portugal*. The last sentence reveals the change in the lexical semantics of *win*. The same is true of *admit*, as in *Teachers admit this exhibition for free*, on a museum notice board advertising a special exhibit.

Inflection is poorly developed. This is not to say that plural marking and verb agreement are completely lacking. More commonly, they are used apparently at random and may occur in unexpected places (recorded telephone message in Singapore: *This transfer will takes about five seconds*). The complex verbal morphology associated with aspectual meanings suffers the same fate. However, differences in the aspect system are often due to underlying systematic differences in the way aspectual meanings are expressed.

Grammar

In the literature, grammatical description of the nonnative varieties, with the possible exception of Singapore English, is not as extensive and detailed as that for English-lexified pidgins and creoles. Nevertheless, from the available descriptions it is possible to appreciate the structural diversity within them. Topic prominence is a significant typological change that has variable structural instantiation among the extant nonnative varieties. The topic structures of Singapore English are as extensive as those in Chinese, its main substrate language. A typical example is *everything also want*, the title of a local comic strip. Here, we see Chinese influence in the fronted topic *everything*, in the adverb *also*, which reinforces the meaning of the quantifier *everything*, and in the missing subject. Related to the topic structure is the novel conditional construction in which the protasis is not introduced by *if*. In *Don't want egg, please inform first*, the protasis is interpreted as the topic that specifies the condition for the apodosis. Typologically or parametrically related syntactic properties tend to cluster in substrate transfer. Other nonnative varieties also allow missing arguments, but they do not have the same range of topic structures (Cheshire, 1991; Baumgardner 1996).

Another significant, and often substrate-driven, change concerns the aspectual system, which varies across the nonnative varieties (Platt *et al.*, 1984). In Singapore English, the perfective aspect is expressed by *already* (*I wash my hands already*), which occurs predominantly in clause-final position. Careful analysis reveals a subtle yet systematic difference between *already* and the past tense or perfect of native English. While *I wash my hands already* may be rendered as *I washed* (or *have washed*) *my hands*, *the wall white*

already means that the wall is white, not that the wall was or has been white. This use of *already* is consistent with the perfective aspect of Chinese.

Register

Nonnative varieties of English do not have an accepted written form, unlike some English-lexified pidgins and creoles, such as Tok Pisin. They are used as a vernacular for informal occasions and have yet to develop a full repertoire of registers – linguistic styles associated with context of use. Newspapers such as *The New Straits Times* (Malaysia), *The Straits Times* (Singapore), and *The Hindu* (India) use native English in their stories, which may contain linguistically trivial neologisms characteristic of the local cultural milieu. Literary works are also written in native English; nonnative varieties are used in the speech of characters as an indexical marker of their low social and educational standing (Talib, 2002). The thin repertoire of registers of a typical nonnative variety is correlated with its limited grammatical resources, its historical roles, and its present sociolinguistic status.

The lack of registrational variation is supported by available corpus evidence. Table 2 displays the frequencies of *already* in the spoken and written registers of Singapore English (SIN) and British English (GB). The data are culled from the International Corpus of English (Greenbaum, 1996). (See Table 3.)

There is no difference in the written register. Differences emerge only in the spoken register of Singapore English, especially in the sentence-final position. The corpus profile suggests a clear registrational division of labor: substrate-driven grammatical innovations are used in informal contexts and avoided in formal contexts.

Stigma and Grammatical Growth

One reason for the underdeveloped state of the nonnative variety is the lack of prestige in the adoptive speech community. Even in places such as Singapore, where English is the *de facto* national language and the local accent is increasingly seen as a marker of the Singaporean identity (Ooi, 2001), grammatical features that deviate from native English are stigmatized

Table 2 Counts of *already* in private conversation and writing, normalized to 1000 words of text

| | Medial position | | Final position | |
|------------------|-----------------|------|----------------|------|
| | SIN | GB | SIN | GB |
| Spoken register | 0.42 | 0.18 | 0.98 | 0.04 |
| Written register | 0.39 | 0.37 | 0.02 | 0.02 |

Note. SIN, Singapore English; GB, British English.

Table 3 Examples of *already* from the Singaporean component of the International Corpus of English

| | |
|-------------------------|---|
| With dynamic predicates | |
| 1. | Maybe she increase the price <i>already</i> |
| 2. | I told you about it <i>already</i> remember |
| With stative predicates | |
| 3. | It's like kind of oldish <i>already</i> |
| 4. | Her hand better <i>already</i> |
| Habitual states | |
| 5. | Nowadays I switch to Mandarin <i>already</i> |
| 6. | I think I am quite used to it <i>already</i> |
| With negatives | |
| 7. | By the time you eat nuh not nice <i>already</i> |
| 8. | Aiyah I cannot remember <i>already</i> |
| In coordinate sentences | |
| 9. | When I was in sec one I noticed him <i>already</i> |
| 10. | If reject then she wouldn't get her PP <i>already</i> |

and frequently targeted for eradication in government-sponsored *Speak Good English* movements. Stigmatization has serious consequences for the non-native variety. Not only do stigmatized features face individual and institutional resistance, they are also slow to stabilize for eventual codification (Bao, 2003). Nonnative varieties need to overcome stigma, reduce internal variation, and expand linguistic resources before they are able to function in wider communicative domains. Against the international prestige and dominance of native English, this is no easy task.

Theoretical Approaches

The bulk of the literature on nonnative varieties of English is devoted to sociolinguistic issues arising from the global spread of English, among them identity, ownership, standardization, and English pedagogy (Quirk and Widdowson, 1985; Cheshire, 1991; Kachru, 1992; Fishman *et al.*, 1996; Görlach, 2002). The cause of grammatical restructuring is also the subject of intense study and lively debate, especially among scholars of pidgins and creoles. Some scholars treat all varieties of English as adaptations to their environment, so the tripartite division – native, nonnative, and pidgin and creole – has little theoretical significance (Mufwene, 1994).

Among the many factors that are involved in grammatical restructuring, we can list linguistic universals, markedness conventions, internal drift, and language acquisition. Also crucial is the role of the languages in the contact ecology, especially the linguistic substratum. The continued presence of indigenous languages in the speech community of a nonnative variety gives added importance to substrate transfer as a prime mover of grammatical restructuring (Lefebvre, 1998). At the same time, native English exerts strong normative pressure. The antagonistic

forces on the grammar of the nonnative variety cannot be resolved purely on linguistic grounds. Grammatical restructuring is a composite and complex process, and no single mechanism is solely responsible. For recent summaries of the field, see Thomason (2001) and Winford (2003).

Bibliography

- Bao Z (2003). 'Social stigma and grammatical autonomy in non-native varieties of English.' *Language in Society* 32, 23–46.
- Baumgardner R J (ed.) (1996). *South Asian Englishes: Structure, use, and users*. Urbana: University of Illinois Press.
- Cheshire J (ed.) (1991). *English around the world: Sociolinguistic perspectives*. Cambridge: Cambridge University Press.
- Fishman J A, Conrad A W & Rubal-Loez A (eds.) (1996). *Post-imperial English: Status change in former British and American colonies, 1940–1990*. Berlin: Mouton de Gruyter.
- Görlach M (2002). *Still more Englishes*. Amsterdam: John Benjamins.
- Greenbaum S (ed.) (1996). *Comparing English worldwide: The international corpus of English*. Oxford: Clarendon.
- Gupta A F (1994). *The step-tongue: Children's English in Singapore*. Clevedon: Multilingual Matters.
- Kachru B B (ed.) (1992). *The other tongue: English across cultures* (2nd edn.). Urbana: University of Illinois Press.
- Kwan-Terry A (ed.) (1991). *Child language development in Singapore and Malaysia*. Singapore: Singapore University Press.
- Lefebvre C (1998). *Creole genesis and the acquisition of grammar: The case of Haitian creole*. Cambridge: Cambridge University Press.
- Mufwene S (1994). 'New Englishes and criteria for naming them.' *World Englishes* 13, 21–31.
- Ooi V (ed.) (2001). *Evolving identities: The English language in Singapore and Malaysia*. Singapore: Times Academic Press.
- Platt J, Weber H & Ho M L (1984). *The new Englishes*. London: Routledge.
- Quirk R & Widdowson H G (eds.) (1985). *English in the world: Teaching and learning the language and literatures*. Cambridge: Cambridge University Press.
- Schneider E W (2003). 'The dynamics of new Englishes: From identity construction to dialect birth.' *Language* 79, 233–281.
- Talib I (2002). *The language of postcolonial literatures: An introduction*. London: Routledge.
- Thomason S G (2001). *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason S G & Kaufman T (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Winford D (2003). *An introduction to contact linguistics*. Malden, MA: Blackwell.

English: World Englishes

B Kachru, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

The conceptualization of the term ‘world Englishes’ is within a ‘socially realistic’ approach to language study (see, e.g., B. Kachru, 1981). The first linguist who, in a rather indirect way, provided such insight about what is now termed world Englishes was John Rupert Firth (1890–1960), the first holder of the chair of general linguistics at London University. In 1956, after his extensive experience in South Asia, Firth (1956: 97) observed:

English is an international language in the Commonwealth, the Colonies and in America. International in the sense that English serves the American way of life and might be called American, it serves the Indian way of life and has recently been declared an Indian language within the framework of the federal constitution. In another sense, it is international not only in Europe but in Asia and Africa, and serves various African ways of life and is increasingly the all-Asian language of politics. Secondly, and I say ‘secondly’ advisedly, English is the key to what is described in a common cliché ‘the British way of life.’

This observation, made over a half-century ago, exemplifies the linguistic pragmatism and social and functional realities of the English language in world context. Firth’s earlier observations have been addressed in much more detail in the following years by a variety of theoretical and methodological frameworks (for a perceptive historical review, see Bolton, 2004).

Spread and Stratification

The cross-cultural and cross-linguistic diffusion of English may be viewed in terms of three phases. The first phase was initiated in the British Isles in 1535 when the Act of Union annexed Wales to England. The linguistic implications of this Act were far reaching, as outlined by Edwards (1993: 108):

The most damaging section of the Act of Union, as far as the Welsh language was concerned and thus a significant element in its collective consciousness, was its emphasis on English as the language of preferment. English became essential for success. It specified “no personne or personnes that use the Welsshe speche or langage shall have or enjoy any manner of office or fees within the Realme of Onglonde Wales or other the Kinges dominions and exercise the speche or langage of Engliche.

In 1603, Scotland also came under British rule, and with this territorial expansion, King James VI became

King James I of England. The expansion continued, and in 1707 yet another non-English speaking region, Ireland and its indigenous languages of Celtic and Gaelic, were subsumed. This phase of expansion was notable for the consolidation of the dominance of English in the British Isles.

It was in the second phase of the diffusion that the diasporic varieties of English were transplanted across continents, notably to North America, Australia, Canada, and New Zealand. This phase involved a significant movement of native-English-speaking populations to new social, linguistic and cultural contexts. Although in total numbers this relocated population was limited, these groups, for example in Australia and New Zealand, developed influential and powerful English-using communities. As time passed, various strategies of educational planning, proselytization, and trading in the language were used to initiate – and increase – bilingualism in English.

The third phase of diasporic expansion introduced English into Asia and Africa. In contrast to the second phase, it brought English into contact with genetically and culturally unrelated languages in far-flung parts of the world. This diaspora provided a new ecology and, for the teaching of English, unprecedented challenges in terms of language contact, cultural contexts, norms, identities, and methodologies. Those challenges continue to confront the professionals in the new millennium.

This diasporic expansion laid the foundations for the use of the English language as cultural ammunition in all these territories and resulted in several indigenous varieties. The reactions to these transplanted varieties and their historical, social, educational, ideational, and cultural implications have ultimately resulted in the most articulate critical debates – both of agony and ecstasy (for further references, see B. Kachru, 1996).

The characterization of the stratification and functions of world Englishes within theoretical and pragmatic frameworks received a further stimulus in the 1970s. It was John Lyons (1978: xvi) who pointed out the parallels “between Labov’s approach to linguistics and that of the ‘British’ school, which draws its inspiration from J. R. Firth.” The ‘socially realistic’ paradigms – mixed with the activism of their proponents – resulted in consideration of linguistic diversity within Englishes as an integral part of social interaction and contextual realities (see B. Kachru, 1981).

Several schemas have been presented to characterize the diffusion of English and its global presence (see McArthur, 1993). One such model that has been

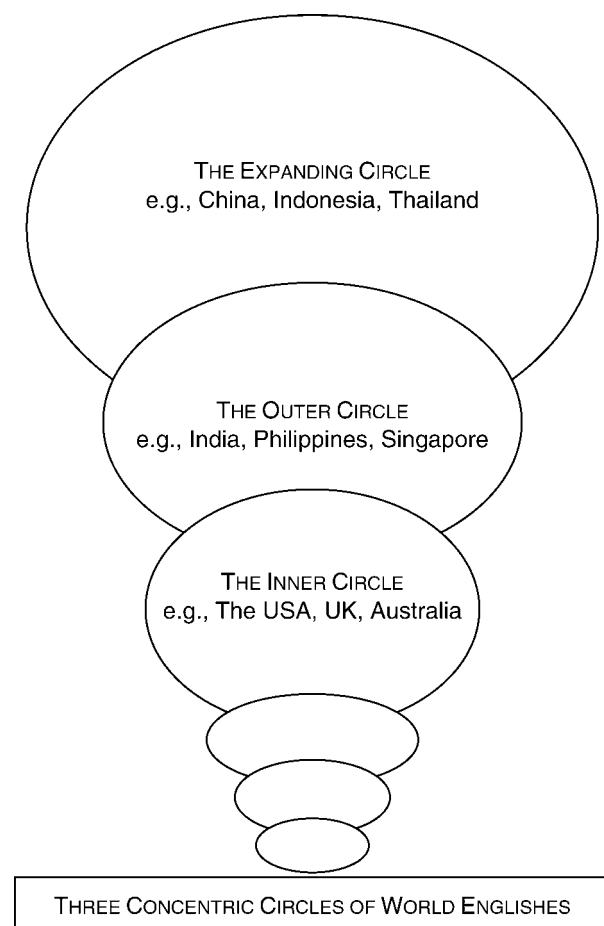


Figure 1 Three concentric circles of World Englishes.

adopted in several studies since the 1980s, the Concentric Circles model (Figure 1), is discussed below in detail.

Concentric Circle Model

The concentric circles representation of the spread of English, proposed in 1985, is more than mere heuristic metaphor for schematizing the spread of English. This representation provides a schema for the contextualization of world varieties of English and their historical, political, sociolinguistic, and literary contexts. The characterization of world Englishes is primarily based on the following factors:

- the history of the types of spread and motivation for the location of the language
- patterns of acquisition
- societal depth of the language in terms of its users, and the range of functions that are assigned to the English medium at various levels in the language policies of a nation (e.g., in administration, education, and literacy)

- functional acculturation of the English language within the local culture and societies and its nativization in the society and its literary culture

The term ‘nativization’ refers to the formal and functional changes the language undergoes at various linguistic levels (e.g., phonetic, lexical, syntactic, discursual, speech acts, literary creativity). In other words, the diffusion of English over centuries involves geographical expansion into regions of the world that had distinct physical realities and social, cultural and linguistic identities. It is in such contexts that the English language acquired ‘functional nativeness.’ It is the extent of functional nativeness in terms of the range and depth of English in a society that determines its impact. The more such functions of English increase in a speech community, the more local identities the variety acquires.

The three circles are not static, but dynamic and changing. The dynamics of the English language in terms of its status, functions, and attitudes toward it are well documented in the case studies of, for example, Bangladesh, Sri Lanka, Malaysia, Indonesia, and even in several Francophone countries.

In historical terms then, the Inner Circle primarily, but not exclusively, comprises the L1 speakers of varieties of English: It is this circle, (e.g., Britain, United States, Canada, Australia, and New Zealand), that provided the springboard for transplanting the language in other parts of the globe. The Outer Circle includes the major Anglophone countries of Africa and Asia, including India, Nigeria, the Philippines, Singapore, and South Africa. The Expanding Circle includes China, Taiwan, Korea, and Saudi Arabia (for the dynamic nature of this circle, see Berns, 2005).

The three circles model, as McArthur (1993: 334) suggested, represents “the democratization of attitudes to English everywhere in the globe.” In his view,

[T]his is a more dynamic model than the standard version, and allows for all manners of shadings and overlaps among the circles. Although ‘inner’ and ‘outer’ still suggest – inevitably – a historical priority and the attitudes that go with it, the metaphor of ripples in a pond suggests mobility and flux and implies that a history is in the making.

World Englishes Speech Communities

The earlier canonical definitions of the concept of ‘speech communities’ do not capture the pragmatic and functional global realities of the English language. Consider, for example, the restricted definition of the term provided by Bloomfield (1933: 42): “a group of people who interact by means of speech.” On the other hand, in Hymes’s view (1974: 47–51),

a speech community is “a social, rather than linguistic entity” that shares “knowledge of rules for the conduct and interpretation of speech.” The views of Firth (1957: 191) contrasted with that of sociologists and anthropologists, when he wrote:

The study of linguistic institutions is thus more specific and positive and on the whole less speculative than the sociological study of societies. Sociologists and social anthropologists are much bolder than linguists in what they find it possible to state in general human terms. To what lengths sociological abstractions can be extended is well-exemplified in Pareto's theory of residues and derivations.

However, Firth also emphasized that a monolithic description of language does not convey the socially and contextually insightful characteristics of language. In his provocative way, Firth (1957: 29) wrote that the “unity of language is the most fugitive of all unities whether it be historical, geographical, national, or personal. There is no such thing as *une langue une* and there has never been.”

The English-using speech communities involve multiple – and often complex – historical, ideational, functional, and attitudinal contexts. In the global context, these fast-growing communities demonstrate varying degrees of competence in the language and its uses in terms of the range of functions and hybridization. These speech communities are primarily of the following types.

- Monolingual users of the language whose one and only language of communication is a variety of English; for example, a large portion of the inhabitants of the United Kingdom, United States, Australia, and Canada. In these countries, too, the number of bilingual users or non-English-using immigrant populations representing multiple languages from Asia, Africa, the Caribbean, Latin America, and Europe is fast increasing.
- Bilingual users of English who acquire English as an *additional* language for communication in those domains of function in which their L1 is not used or is not considered functionally appropriate to use.
- Multilingual users in whose verbal repertoire English is yet another code of communication, and language-shift and alternation are a normal communicative strategy. This phenomenon has been well documented with reference to multiple Anglophone English-using speech communities in Africa, Asia, Europe, and in the United States and United Kingdom.
- Bidialectal speakers and those whose L1 dialect has not attained functional and attitudinal recognition, as is the case of Ebonics (African-American English) or Spanglish in the United States.

Table 1 The statistics of World Englishes

| Society | Approximate population (million) | Percentage of L1/ L2 English users | Approximate totals (million) |
|-------------------------|----------------------------------|------------------------------------|------------------------------|
| INNER CIRCLE | | | |
| United States | 293 | | |
| United Kingdom | 59 | | |
| Canada | 32 | | |
| Australia | 20 | | |
| New Zealand | 4 | | |
| OUTER CIRCLE | | | |
| India | 1000 | 33 | 330 |
| Philippines | 86 | 56 | 48 |
| Pakistan | 159 | 11 | 17 |
| Malaysia | 24 | 32 | 8 |
| Bangladesh | 141 | 5 | 7 |
| Hong Kong | 7 | 35 | 2 |
| Singapore | 4 | 50 | 2 |
| Sri Lanka | 20 | 10 | 2 |
| EXPANDING CIRCLE | | | |
| China | 1300 | 18 | 234 |
| Japan | 127 | 33 | 42 |
| Indonesia | 238 | 5 | 12 |
| Thailand | 60 | 10 | 6 |
| South Korea | 49 | 9 | 4 |
| Vietnam | 83 | 5 | 4 |
| Myanmar | 43 | 5 | 2 |
| Taiwan | 22 | 10 | 2 |
| Cambodia | 13 | 5 | 0.6 |
| Laos | 6 | 5 | 0.3 |

The above figures are ‘guesstimates’ based on various published resources.

One major factor that distinguishes the international profile of English from that of other languages of wider communication is that it has more users now who have acquired it as an L2, L3, or L4 in their language repertoire (see Table 1).

Process of Nativization and Englishization

The speech communities of English in the Outer Circle use institutionalized varieties of English, which have the following characteristics:

- recognition of English in the overall language policy of the English-using nation (e.g., India, Nigeria, Singapore)
- an extended tradition of contact literatures in English that are recognized as part of the national literatures.
- social penetration of the language that has resulted in several social, ethnic or functional subvarieties (e.g., Singlish, Basilect, Bazaar English, Tanglish)
- distinct linguistic exponents of the process of nativization at various levels

- an extended range of localized genres and registers
- Englishized varieties of local languages, some of which may have even acquired distinct names (e.g., *Hinglish* or *Hindlish* of India)
- acculturation of the English language for articulating local social, cultural, and religious identities.

The process of nativization is one major linguistic dimension of acculturation of world Englishes: this acculturation is evident in Anglophone Asian and African functionally localized contexts.

The two processes of nativization and Englishization are Janus-like, two faced. One face reflects the impact of contact and convergence with other languages – Asian and African – at various linguistic levels. The second face shows the impact that the English language and literature have on other languages and literatures of the world. Englishization is not restricted to phonology, grammar, and lexis, but can have a deep impact on discourse, registers, styles, and literary genres in contact literatures in Englishes (see, e.g., Thumboo, 1992; Dissanayake, 1997; Y. Kachru, 1997; B. Kachru, 2003; Y. Kachru, 2003; B. Kachru, 2005).

The process of Englishization is evident in three major geographical regions associated with the spread of English:

1. Traditional regions of cultural and literary contact in which a number of cognate languages of English are used (e.g., in Western Europe and parts of Eastern Europe)
2. Anglophone, geographically noncontiguous with English, regions of the Raj, which include the Outer Circle of English, and have, in a genetic sense, unrelated or not-closely related languages (e.g., parts of Africa and Asia)
3. Expanding Circle, which includes the rest of the world (e.g., Japan, China, the Middle East, and Latin America).

In defining the nativeness of varieties of Englishes, a distinction may be made between genetic and functional nativeness. Genetic nativeness refers to the historical relationship of the languages in contact, and functional nativeness to the domains of use of English, the range and depth in social penetration, and the resultant acculturation. A profile of the functional nativeness of a variety of English includes these factors:

- sociolinguistic status of a variety in its transplanted context
- range of functional domains in which a variety is used
- creative processes used to construct localized identities

- linguistic exponents of acculturation
- types of cross-over contributing to canons of creativity
- attitude-specifying labels used for the variety of English.

The second diaspora of English has raised a variety of questions that are unique to transplanted Englishes and continue to be debated in the literature (see, e.g., B. Kachru, 1988, 1996; Mufwene, 2001; B. Kachru, 2005).

Models of Description

The canonical models of English continue to be viewed in terms of privileged British and American varieties of the language. The theoretical, methodological, and ideational issues raised by such an attitude have been extensively – and passionately – articulated in the literature in recent years. This debate has acquired a prominent position in the conceptualization of world Englishes. There are essentially three types of speech fellowships of world Englishes: (1) those that are canonically considered privileged and norm-providing – the Inner Circle, (2) those that have functionally acquired the status of norms and are pragmatically relevant in their sociolinguistic context – the Outer Circle, and (3) those that in many respects continue to be attitudinally dependent on external norms, primarily from the Inner Circle – the Expanding Circle (see Berns, 2005).

Conceptual Myths

The articulation of the following six myths in the conceptualization, methodology, and pedagogy of world Englishes has resulted in the ‘paradigms of marginalization.’ These paradigms are essentially based on age-old following fallacies:

1. World Englishes in Anglophone Asia and Africa are acquired and used to interact with canonical ‘native’ speakers of English.
2. World Englishes are acquired to learn the Judeo-Christian traditions as articulated in American and British cultural and literary values and traditions.
3. The Inner Circle Englishes are primary and standard ‘model providers’ for teaching and acquiring communicative competence in the language.
4. Conceptually, all varieties of world Englishes in Outer Circle are essentially *deficit* or *interlanguage* varieties.

5. Historically it is the Inner Circle that has provided – or should provide – models and standards for ELT pedagogy, creativity, and canonicity of Englishes across Anglophone regions and cultures.
6. The arms of codification of the English language, established – and imposed – by agencies of the Inner Circle, should ideally control the variation and diversity in world Englishes (see the Quirk-Kachru controversy, discussed in detail in Tickoo, 1991).

Constructing Identities of Englishes

The controversial modifiers of the term ‘English’ that are frequently used to characterize the post-colonial diffusion and stabilization of the English language across cultures and languages include ‘new Englishes’ and ‘international,’ ‘global,’ or ‘world English.’

The term ‘New Englishes’ was primarily – though not exclusively – used for the institutionalized varieties in the Outer Circle. All the ‘new’ varieties are transplanted (diaspora) varieties that have a presence on almost every continent. However, the use of the modifier ‘new’ for such Englishes is a misnomer – historically, contextually, and in terms of their acquisition, as some of them pre-date some Inner Circle varieties.

The conceptualization of ‘world Englishes’ (and not ‘world English’), in the sense in which it is used here, goes back to the 1960s. Its formal and functional implications were discussed in 1978 in two independently organized conferences in the United States: one at the East-West Center at Honolulu, (April 1–15) and the other at the University of Illinois at Urbana-Champaign (June 30–July 2). The Honolulu conference concluded with a statement and agenda for the future recognizing that “English used as an international and auxiliary language has led to the emergence of sharp and important distinction between the uses of English for international (i.e., external) and intranational (i.e., internal) purposes.”

In addition to this distinction between the uses of English-using speech communities, the statement further distinguished between “those countries (e.g. Japan) whose requirements focus upon international comprehensibility and those countries (e.g. India) which in addition must take account of English as it is used for their own national purposes.” The Honolulu conference also expressed concern that “[s]o far as we know, no organization exists that takes into account of any language in the light of this fundamental distinction.”

The University of Illinois conference, in contrast, “broke the traditional pattern of such deliberations:

no inconvenient question was swept under the rug. The professionals, both linguists and literary scholars, and native and non-native users of English, had frank and stimulating discussions” (Kachru, 1997: 210).

The scholars present, almost all from Anglophone countries – including, Africa and Asia, as equal partners – discussed with refreshingly fresh perspectives the sociolinguistic and linguistic profile of each English-using country in terms of the functional range of their varieties of Englishes and the social depth of the penetration of the language. What emerged were fascinating worldwide profiles of nativization and acculturation of world Englishes and construction of their identities, attitudes, and functions. It was through such discussions that a socially realistic and pragmatically appropriate preliminary framework developed.

This socially realistic framework represents the formal and functional variations, divergent sociolinguistic contexts, and histories of world Englishes. It is through such contextual insights that the bilinguals’ creativity, at various levels, acquires a social and functional meaning. The concept underscores the ‘WE-ness’ of the medium, its distinct nativeness determined in cultural, linguistic, and ideological contexts of Anglophone communities. Such cross-cultural functions of the medium acquire their own semantic signals in which the traditional dichotomies and frameworks demand alternative approaches. There is recognition of the fact that different methodologies may be needed (e.g., literary, linguistic, and pedagogical) to capture and construct the altered identities represented in the medium in Englishes of the world. The pluralization of the canonical term ‘English’ does not suggest ‘divisiveness’ in the English-using speech communities, but rather the recognition of a unique functional reality of the language: the diversity of the medium and its assimilative qualities in multiple pluralistic, linguistic, and cultural contexts.

These functional, contextual, and ideational connotations – and realism – are absent, as mentioned above, in such terms as ‘international English,’ ‘global English,’ or ‘world English.’ The term ‘international’ is misleading in more than one sense: it signifies an international English in terms of acceptance, proficiency, function, norms, and intelligibility. These presuppositions are far from the real world of Englishes in the world contexts.

The other concept currently presented to represent the global – or some times a regional – medium is ‘lingua franca English.’ This term was originally restricted to the intermediary contact language

(*Vermittlungssprach*) used by the Arabs and the Turks as a maritime jargon in the Levant. It primarily signifies a language of commerce (e.g., Italian around the Adriatic Sea). Each variety in the Outer Circle, as in the Expanding Circle, has its subvarieties in terms of functional connotations, domains, and attitudes toward localized varieties of Englishes and their cross cultural and cross-linguistic communications. Yet, the Inner Circle has made no serious efforts – socially, methodologically, or pedagogically – to recognize their status and currency.

One often-quoted interpretation of the concept of world Englishes was provided by McArthur (1993: 334) when he referred to the logo-acronym of the journal *World Englishes* (which started in 1984), which “serves to indicate that there is a club of equals here.” In this interpretation, the emphasis is on “the democratization of attitudes to English everywhere on the globe,” and it, as McArthur perceptively pointed out, dissolves the trinity of ENL, ESL, and EFL nations.

The linguistic, cultural, canonical, and literary implications of the diffusion of English beyond the Inner Circle are discussed in, for example, Dissanayake (1997), Thumboo (1992), and B. Kachru (1988, 2005).

World Englishes and Conceptual Frameworks

The theoretical, methodological, and ideological questions related to world Englishes go beyond language pedagogy, which was the primary concern before the 1950s. In the post-1960s period, several sacred linguistic cows of theoretical and applied linguistics as applied to world Englishes have been under attack as a consequence of several developments: insights gained by critical sociolinguistic paradigms, the articulation of identities with the language, and altered dynamics of the functions of English in post-colonial linguistic and cultural contexts. One thinks of, for example, the earlier theoretical and methodological emphasis given to such concepts as interference, interlanguage, and fossilization in paradigms of language acquisition. There was very little – if any – awareness of the pluricentricity of Englishes in the Outer Circle or of developing literary and cultural canons and nativized registers and genres in world Englishes in Africa, Asia, and the diasporic writers in the Inner Circle. After the 1960s, a vibrant debate started about several pedagogical issues, such as idealized models for the codification of English, the cross-linguistic claims made for teaching methods and methodologies, and English-language teaching

materials developed, published, and often exported by the English-language teaching ‘experts.’

Two often-articulated descriptive and prescriptive questions – specifically about the Outer Circle varieties – are the following: what criteria may be used to determine a difference between an error (or mistake) and an innovation? And, what variables determine intelligibility for varieties of world Englishes across cultures and languages?

In his extensive empirical research on the latter topic, Smith (1992) viewed intelligibility in a pragmatic communicative context by making a distinction among intelligibility (word utterance recognition), comprehensibility (word utterance meaning [locutionary force]), and interpretability (meaning behind the word/utterance [illocutionary force]). Smith and Nelson (1985) have also discussed some issues that should be on the agenda of any researcher studying intelligibility.

Literary Creativity, Canonicity, and World Englishes

The creative linguistic processes that result in competence in two or more languages are termed ‘contact’ or ‘interference’ varieties. The underlying process in the construction of contact literary texts is that of hybridization, as reflected in bilinguals’ (or multilinguals’) creativity. Such texts are designed with a blend of linguistic features from two or more – related or unrelated – languages. The concept of ‘contact literatures’ thus brings to the English language the multilingual and multicultural contexts of, for example, Africa and Asia. These varieties of English have acquired stable characteristics in terms of pronunciation, grammar, lexis, discoursal, and stylistic strategies. These traditions are often blended with local subvarieties of English, (e.g., Nigerian Pidgin in Nigerian English, Singlish in Singapore English, Bazaar or Babu English in Indian English). In such contact situations, the English language is a medium that has been, pragmatically and contextually, localized to adapt to – and to represent, as elegantly claimed by such writers as Raja Rao and Salman Rushdie (India), Chinua Achebe and Wole Soyinka (Nigeria), Edwin Thumboo and Catherine Lim (Singapore), and F. Sionil Jose (the Philippines).

It is ‘contact’ at various levels (linguistic, social, and cultural) and the resultant nativization that contact literatures represent in literary and cultural canons that are distinct from the Judeo-Christian canons. These processes thus ultimately result in, say, the Africanization or Asianization of world Englishes.

The term ‘interference varieties’ – though attitudinally loaded – is yet another label to conceptualize the contact varieties of English and the bilinguals’ creativity. The interference varieties, as Quirk *et al.* (1985: 27–28) recognized are

so widespread in a community and of such long standing that they may be thought stable and adequate enough to be institutionalized and hence to be regarded as varieties of English in their own right rather than stages on the way to more native-like English.

All such varieties, as shown in numerous studies, have formal and functional identificational features that represent the linguistic processes at various levels: grammar, phonetics, lexis, discourse, speech acts, genres, and indeed literary creativity (see Smith and Forman, 1997; B. Kachru *et al.*, 2005). These studies are of three types: variety-specific (e.g., Indian English, Singapore English, Nigerian English), area-specific (e.g., South Asian English, West African English, Southeast Asian English), or of larger geographical regions in terms of linguistic, literary, and sociolinguistic areas (Africanization or Asianization).

The study of bilinguals’ creativity demands recognition that the institutionalized Englishes have an educated variety and a cline of subvarieties, that writers in contact literatures engage in ‘lectal mixing,’ and that in such texts there are style shifts related to the underlying context of situation. In contextual terms, style shifts result in the construction of altered discourse strategies, speech acts, and registers. In discussing such creativity in world Englishes, Thumboo (1992: 270) argued the following:

This challenge confronts almost every bi-or multilingual writer. His bilingualism is one of three broad types – proficient, powerful, or limited; his position in this cline is not static, because quite often one language gains dominance. A bilingual person has at least two language universes, and each language works with its own linguistic circuits. How the two associate depends on whether the languages as neighbors inhabit the same space and time and can bend to serve creative purposes.

There is thus multicanonicity in world Englishes that blends two or more ‘language universes’ in their creativity; the interlocutors in Englishes have a variety of linguistic, cultural, social, and literary traditions – a speaker of a Bantu language interacting with a speaker of Japanese, a Taiwanese with an Indian, and so on. The traditional and much discussed and canonical ‘native speaker’ may rarely be part of such interactions in Englishes. The linguistic

historical analogues that come to mind – though not necessarily parallel to world Englishes – are that of Latin in medieval Europe (Kahane and Kahane, 1979) and of Sanskrit in traditional South Asia and beyond.

The Pandora’s Box and World Englishes

The shared strands of current debates on world Englishes include the following seven major contextually, attitudinally, pedagogically, and linguistically relevant issues:

1. The demythologization of conventional sacred cows model initiated and nurtured by the Inner Circle constructs of English (see B. Kachru, 1988; Quirk, 1988)
2. The ecologies of multilingual Englishes, specifically in the contexts of Africanization and Asianization of Englishes (see Mufwene, 2001; B. Kachru, 2005)
3. The increasing expression of bilinguals’ creativity in the Outer Circle and its implications on traditional canons and canonicity
4. The theoretical, methodological, and pedagogical implications of the increasing depth and range of Englishes (B. Kachru, 2001)
5. The issues of intelligibility in cross-cultural and cross-linguistic communication
6. The evaluation of ethical practices related to forms, functions, and pedagogy (see, e.g., Baumgardner and Brown, 2003; Dhillon, 2003)
7. The motivation of power and politics and the role of initiators of arms of control (Phillipson, 1992).

Bibliography

- Baumgardner R J & Brown K (2003). ‘World Englishes and pedagogy.’ *World Englishes* 22(3), 245–251.
- Berns M (2005). ‘Expanding on the expanding circle: where do we go from here?’ *World Englishes* 24(1), 85–93.
- Bloomfield L (1933). *Language*. New York: Holt, Rinehart and Winston.
- Bolton K (2004). ‘World Englishes.’ In Davies A & Elder C (eds.) *The handbook of applied linguistics*. Oxford: Blackwell Publishing. 367–396.
- Crystal D (1995). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Dhillon P A (2003). ‘Ethics in language: the case from world Englishes.’ *World Englishes* 22(3), 217–226.
- Dissanayake W (1997). ‘Cultural studies and world Englishes.’ In Smith & Forman (eds.). 126–145.

- Edwards G D (1993). 'The response of education to the native language of Wales.' In Khoo R, Kreher U & Wong R (eds.) *Languages in contact in a multilingual society: implications for language teaching and learning*. Singapore: SEAMEO Regional Language Centre. 107–128.
- Firth J R (1956). 'Descriptive linguistics and the study of English.' In Palmer F R (ed.) *Selected papers of J. R. Firth 1952–59*. Bloomington, IN: Indiana University Press. 1952–1959.
- Firth J R (1957). *Papers in linguistics 1934–51*. London: Oxford University Press.
- Firth J R (1959). 'The treatment of language in general linguistics.' In Palmer F R (ed.) *Selected papers of J. R. Firth 1952–59*. London: Longmans.
- Hymes D (1974). *Foundations in sociolinguistics: an ethnographic approach*. Philadelphia: University of Pennsylvania Press.
- Kachru B B (1981). '“Socially realistic linguistics”: the Firthian tradition.' *International Journal of Sociology of Language* 31, 65–89.
- Kachru B B (1985). 'Standards, codification and sociolinguistic realism: the English language in the outer circle.' In Quirk R & Widdowson H G (eds.) *English in the world: teaching and learning the language and literatures*. Cambridge: Cambridge University Press. 11–30.
- Kachru B B (1988). 'The spread of English and sacred linguistic cows.' In Lowenberg P H (ed.) *Language spread and language policy: issues, implications, and case studies*. (GURT 1987). Washington, DC: Georgetown University Press. 207–228.
- Kachru B B (1996). 'World Englishes: agony and ecstasy.' *Journal of Aesthetic Education* 30(2), 135–155.
- Kachru B B (1997). 'World Englishes 2000: resources for research and teaching.' In Smith L E & Foreman M L (eds.) *World Englishes 2000*. Honolulu, HI: University of Hawaii Press. 209–251.
- Kachru B B (2001). 'World Englishes and culture wars.' In Goh R *et al.* (eds.) *Ariels-departures and returns*. Singapore: Oxford University Press. 391–414.
- Kachru B B (2003). 'On nativizing *mantra*: Identity construction in Anglophone Englishes.' In Ahrens R, Parker D, Stierstorfer K & Tam K-K (eds.) *Anglophone cultures in south-east Asia: appropriations, continuities, contexts*. Heidelberg, Germany: Universitätsverlag Winter Heidelberg. 55–72.
- Kachru B B (2005). *Asian Englishes beyond the canon*. Hong Kong: Hong Kong University Press.
- Kachru B B Kachru Y & Nelson C (eds.) (2005). *The handbook of World Englishes*. Oxford: Blackwell.
- Kachru Y (1997). 'Culture and argumentative writing in world Englishes.' In Smith & Forman (eds.). 48–67.
- Kachru Y (2003). 'Conventions of politeness in plural societies.' In Ahrens R, Parker D, Stierstorfer K & Tam K-K (eds.) *Anglophone cultures in south-east Asia: appropriations, continuities, contexts*. Heidelberg, Germany: Universitätsverlag Winter Heidelberg. 39–53.
- Kahane H & Kahane R (1979). 'Decline and survival of Western prestige languages.' *Language* 55, 183–198.
- Lyons J (1978). 'Foreword.' In Labov W (ed.) *Sociolinguistic patterns*. Oxford: Blackwell. xi–xxiii.
- McArthur T (ed.) (1992). *The Oxford companion to the English language*. Oxford: Oxford University Press.
- McArthur T (1993). 'The English language or the English languages?' In Bolton W F & Crystal D (eds.) *The Penguin history of literature*. 10: *The English language*. London: Penguin Books. 323–341.
- Mufwene S (2001). 'New Englishes and norm setting: how critical is the native speaker in linguistics?' In Thumboo E (ed.) *The three circles of English*. Singapore: UniPress. 133–141.
- Phillipson R (1992). *Linguistic imperialism*. Oxford: Oxford University Press.
- Quirk R, Greenbaum S, Leech G & Svartvik J (1985). *A comprehensive grammar of the English language*. London: Longman.
- Quirk R (1988). 'The question of standards in the international use of English.' In Lowenberg P H (ed.) *Language spread and language policy: issues, implications, and case studies*. (GURT 1987). Washington, DC: Georgetown University Press. 229–241.
- Smith L E (1992). 'Spread of English and issues of intelligibility.' In Kachru B B (ed.) *The other tongue: English across cultures*, 2nd. edn. Urbana, IL: University of Illinois Press. 75–90.
- Smith L E & Forman M L (eds.) (1997). *World Englishes 2000*. Honolulu, HI: University of Hawaii Press.
- Smith L E & Nelson C L (1985). 'International intelligibility of English: directions and resources.' *World Englishes* 4(3), 333–342.
- Thumboo E (1985). 'Twin perspectives and multi-ecosystems: tradition for a commonwealth writer.' *World Englishes* 4(2), 213–221.
- Thumboo E (1992). 'The literary dimensions of the spread of English.' In Kachru B B (ed.) *The other tongue: English across cultures*, 2nd. edn. Urbana, IL: University of Illinois Press. 255–282.
- Thumboo E (ed.) (2001). *The three circles of Englishes: language specialists talk about the English language*. Singapore: UniPress.
- Tickoo M L (ed.) (1991). *Languages and standards: issues, attitudes, case studies*. Singapore: South East Asian Ministers of Education Organisation (SEAMEO).

Eskimo–Aleut

A Berge, University of Alaska, Fairbanks, AK, USA

© 2006 Elsevier Ltd. All rights reserved.

The Terms ‘Eskimo’ and ‘Aleut’

The terms ‘Eskimo’ and ‘Aleut’ have unclear origins. ‘Eskimo’ is commonly believed to be derived from a derogatory Algonquian term meaning ‘eaters of raw meat’; it has been replaced with Inuit, meaning ‘people,’ in most of Canada, and the language is referred to variously as Inuinnaqtun, Inuttun, Inuktitut, and by other names. The term ‘Aleut,’ confusingly, was bestowed on Aleut and Yupik people native to the Aleutian Islands and part of the southwestern Alaskan mainland, as well as on their languages, by Russians in the 18th century. The term in use in Alaska today for self-designation of the non-Yupik Aleut is Unangan or Unangas for the people and Unangam Tunuu for the language; however, the term ‘Aleut’ is preferred in Russia. Because there is still no other general term to describe all of the languages and dialects encompassed by the terms ‘Eskimo’ and ‘Aleut,’ and for reasons of linguistic tradition, they are still commonly used within the field of Eskimo–Aleut linguistics.

History and Genetic Relationships

The Eskimo–Aleut language family is spoken from the Siberian coast in the west to Greenland in the east. There are two major branches, Aleut and Eskimo, and Eskimo consists of at least two further subgroups, Yupik and Inuit. One recently extinct language, Sireniki (Sireniki Yupik), may either have been a third branch of Eskimo or one of the Yupik languages.

The Eskimo and Aleut people are thought to have been part of the last large-scale migration from Asia across the Bering land bridge, between 4000 and 6000 years ago. Various attempts have been made to link Eskimo–Aleut with other language families on the Asian continent (e.g., Uhlenbeck, 1935, Swadesh, 1962). While there is little solid linguistic evidence of a genetic relationship between Eskimo–Aleut and other language families, there are strong suggestions of very early contact, particularly with Uralic (for a discussion of possible linguistic affinities and contact, see Fortescue, 1998). The development and differentiation of the Aleut and Eskimo languages and dialects probably took place in Alaska because of the linguistic diversity found on the Alaskan side.

From Alaska, there were several waves of migration down to the Aleutians, west again to Siberia, and

eastward to Greenland. The earliest led to the linguistic split between Aleut and Eskimo, possibly around 4000 years ago, although there is archeological and skeletal evidence to suggest an earlier physical divergence (Laughlin, 1980). The earliest Aleut settlements appear to have been around the Island of Four Mountains, whence there were eastward and westward migrations. The Yupik and Inuit branches of Eskimo must have diverged about 2000 years ago. From their homeland around the Seward Peninsula, Yupik speakers occupied southwestern Alaska and, moving westward across the Bering Strait, reoccupied the Chukchi Peninsula in Siberia. If Sireniki is a separate branch of Eskimo, then it may have split off about 2500 years ago and its origins may be on the Chukchi Peninsula as a result of an earlier back migration. It was gradually displaced by Central Siberian Yupik and neighboring Chukchi (Chukot). The ancestors of the Inuit spread northward and eastward in several waves. The latest migration began about 1000 years ago, and resulted in the rapid spread of the Inuit language, leaving few linguistic traces of previous Eskimo populations. The present dialect differentiation (see Figure 1) is possibly as recent as the past 500 years (see Dorais, 1996).

Historiography of Research

The first systematic linguistic studies of an Eskimo–Aleut language were made in Greenland, beginning in the 18th century. Outside of Greenland, most available materials on the Eskimo–Aleut languages consisted of word lists until well into the 19th century, and even into the 20th century for some of the western Inuit dialects (see *West Greenlandic* and *Inupiaq*). Nevertheless, the Danish scholar Rasmus Rask proposed a genetic relationship between Aleut and Eskimo languages as far back as 1819. His notes were published in 1916 (Thalbitzer, 1916), and his thesis was confirmed through Marsh and Swadesh (1951) and Bergsland (1951). More recently, comparative work has been done by Bergsland (1986, 1989) and Fortescue *et al.* (1994).

Linguistic Characteristics of Eskimo–Aleut

The following features are characteristic of the Eskimo–Aleut language family:

- There are three basic vowels (i, u, a), derived from an original four-vowel system (i, u, a, and schwa, represented by [e] in Yupik, which maintains the four-vowel distinction).

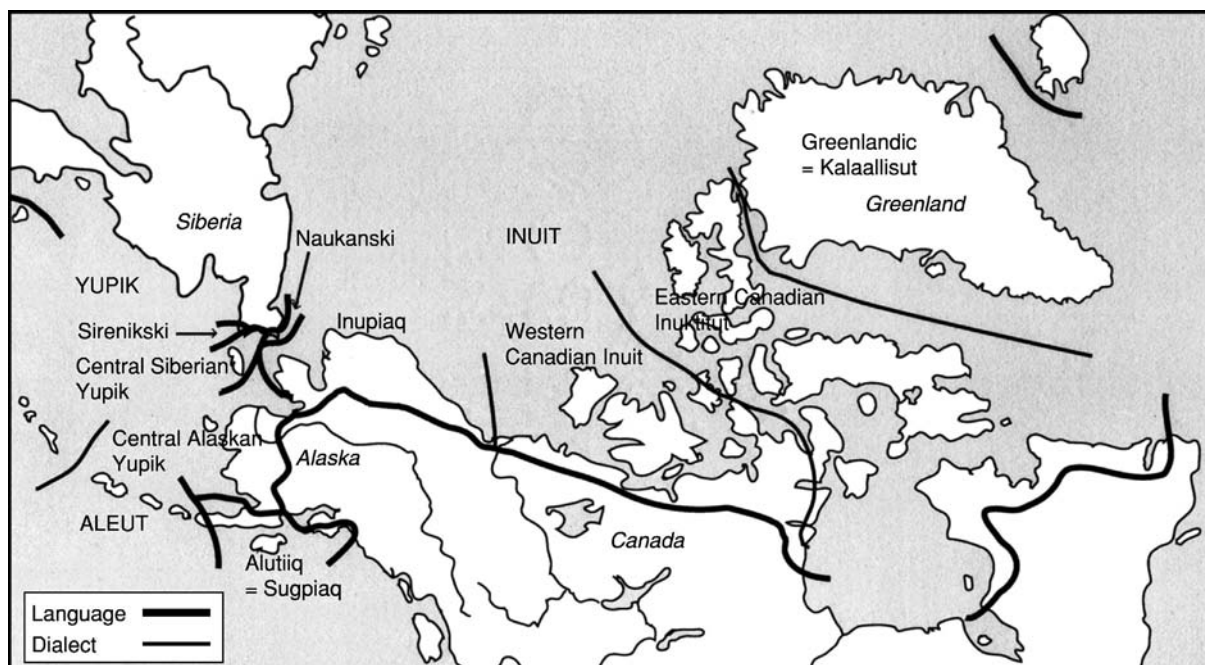


Figure 1 Eskimo-Aleut languages and major dialects. (Adapted from Map 1 in Fortescue *et al.* (1994)).

- Word building is almost exclusively through suffixation, the only exception being an anaphoric prefix *ta-* on demonstrative stems (e.g., pan-Eskimo *una* (DEIC-PRX), *tauna* ‘this one’).
- To a greater or lesser degree within the language family, polysynthesis is the norm: very complex verb structures encode meanings for which other languages need whole sentences. This example is from Uummarmiut Eskimo (from Inuvik):

kivgaluk-niaq-qati-gi-tqik-kuminait-kiga
muskrat-hunt-partner-have.as-again-will.never-
 1SG.3SG.INDIC

‘I will never again have him as a partner to hunt muskrat’ (Lowe, 1985: 18)

- Sentences typically consist of clause chains, in which a series of dependent clauses is headed by an independent clause. In the following simple example from West Greenlandic, there is one indicative clause and two subordinated participial clauses; clause chains can be quite extensive:

irn-i qajartur-tuq
son-his own be out in kayak-3SG.PART
 qinnguar-paa
see through binoculars-3SG.3SG.INDIC
 natsirsu-up sursuk-kaa
hooded seal attack-3sg.3sg.PART

‘he_i saw his son_j through his binoculars being attacked by a hooded seal while in his_j kayak’
 (Fortescue, 1984: 39)

- Word order is typically SOV to a more or less fixed degree (Aleut has essentially fixed word order).

In the previous example, the object clause *irni qajarturtuq* precedes the verb, and the subject of each of the subordinated clauses precedes its respective verb.

- Case marking follows the ergative-absolutive case-marking pattern (with extreme modifications in Aleut), where the subject of intransitive verbs receives the same marking as the object of transitive verbs, namely absolutive case, while the subject of transitive verbs receives a different marking, ergative case.

The Eskimo languages are much more closely related to each other than to Aleut. In addition to the common features listed for Eskimo-Aleut, they share

- certain restrictions on syllable and other phonological structures;
- up to six nongrammatical cases in addition to ergative and absolutive cases (locative, instrumental, ablative, allative, vialis, and equalis);
- transitive and antipassive structures, the choice of which appears to be partially determined by definiteness, as in the Central Alaskan Yup’ik (Central Yupik) examples below:

| | | |
|--------------------------------|--------------------|--------------------------|
| angute-m | neqa | ner-aa |
| <i>man-ERG.SG</i> | <i>fish-ABS.SG</i> | <i>eat-3SG.3SG.INDIC</i> |
| ‘the man eats the fish’ | | |

| | | |
|------------------------------|----------------------|----------------------|
| Angun | neq-mek | ner’-uq |
| <i>man-ABS.SG</i> | <i>fish-INSTR.SG</i> | <i>eat-3SG.INDIC</i> |
| ‘the man eats a fish’ | | |

Aleut

Aleut is a language with two major extant dialects, and at least two other dialects historically attested. Eastern Aleut is spoken east of Amukta Island to the Alaskan Peninsula, as well as on the Pribilof Islands. Atkan, also variously called Western or Central Aleut, is today spoken on Atka Island, and a version of it is spoken on Bering Island. A third dialect, Attuan, is essentially dead; a mixed language known as Copper Island Aleut (Mednyi Aleut), consisting of Attuan stems and Russian inflection, is still spoken on Bering Island. There is scant evidence in very early descriptions of a dialect spoken on the Rat Islands between Attu and Atka, suggesting a dialect continuum in the West.

Characteristic of Aleut are

- its lack of labial stops (although it appears that Copper Island Aleut has developed voiced labial stops), and its aspirated nasals;
- consonant clusters which differ from those in Eskimo in their distribution (e.g., they are permitted word initially, as in *qdinalix* 'to be slippery'), in the combinations of consonants possible (e.g., velar-uvular fricatives), and in their complexity (allowing up to three consonants, as in *chǎxsxǐx̌* 'reef');
- use of independent pronouns, as opposed to pronominal marking on verbs in Eskimo languages:

| | | |
|-------|------------|-----------------------|
| Aleut | txin | achix-ku-qing |
| | <i>you</i> | <i>teach-PRES-1SG</i> |
| | | 'I am teaching you' |

| | |
|-------------|----------------------------------|
| Greenlandic | ilinniar-tip-pakkit |
| | <i>learn-cause-1SG.2SG.INDIC</i> |
| | 'I am teaching you' |

- its typologically unusual agreement system, in which ergative case marking is only used if a transitive object or an object of possession is not overtly expressed:

| | | |
|---------------------|-------------------|----------------------------|
| Piitra-̌x̌ | tayǎgu-̌x̌ | kidu-ku-̌x̌ |
| <i>Peter-ABS.SG</i> | <i>man-ABS.SG</i> | <i>help-PRES-3SG</i> |
| | | 'Peter is helping the man' |

| | |
|---------------------|--------------------------|
| Piitra-m | kidu-ku-u |
| <i>Peter-ERG.SG</i> | <i>help-PRES-3SG.3SG</i> |
| | 'Peter is helping him' |

Through contact with Russian traders and colonizers in the 18th and 19th centuries, modern Aleut has a large proportion of Russian loanwords. As a result of early decimation of the people and later suppression of the language in the schools, the language is severely endangered today, with at most 150 speakers in the

Aleutian Islands, the Pribilof Islands, and Anchorage, and perhaps 10 on Bering Island. With the exception of Atkan, speakers are elderly.

Yupik

The Yupik languages include Naukanski (Naukan Yupik), spoken around East Cape on the Chukchi Peninsula; Central Siberian Yupik, spoken on the Chukchi Peninsula and on St Lawrence island; Central Alaskan Yup'ik, spoken from Norton Sound to Bristol Bay in Alaska; and Sugpiaq (Pacific Gulf Yupik), also known as Sugcestun or Alutiiq, spoken around Prince William Sound, the Alaskan Peninsula, Kodiak and Afognak Islands, and the tip of the Kenai Peninsula. Yupik languages are characterized by

- their retention of a fourth vowel that presumably stems from Proto-Eskimo (cf. Proto-Eskimo **neqe* became Central Alaskan Yup'ik *neqa* 'food,' Iñupiaq [North Alaskan Inupiatun] *niqi* 'food');
- more or less complex effects of intonational stress; in stressed syllables, for example, the vowel is often lengthened (for more on Yupik prosody, see Krauss, 1985).

There are some nonnegligible syntactic differences between the languages, although these have not yet been well described. Siberian Yupik languages have a number of English loan words, from contact with 19th-century whalers, and Alaskan Yupik languages have a large number of Russian loans from 18th- and 19th-century Russian colonization, as well as 20th-century English loans. Most Yupik languages are severely endangered today, with numbers of speakers ranging from 70 (Naukanski) to 1300 (Central Siberian Yupik). The notable exception is Central Alaskan Yup'ik, with about 10 000 speakers (and on the Kenai Peninsula these include children), and with immersion programs in the schools and active production of learning materials.

Sirenikski

Sirenikski is seen as an important link to Proto-Eskimo because of particularly conservative features, such as a retention of consonants between vowels, which were lost in all other Eskimo-Aleut languages (e.g., Proto-Eskimo **ataRuciR*, Sirenikski *ategesegh*, Central Alaskan Yup'ik *atauciq*, Iñupiaq *atausiq* 'one'). It has, however, undergone sound changes quite different from other Eskimo languages. For example, in all nonstressed (essentially, noninitial) syllables (as in the example given above) the vowel changed to schwa. Unfortunately, it was first

discovered and described at the end of the 19th century, when it was already highly moribund; the last speaker died in the year 2000.

Inuit

Inuit is generally described as a language with four distinct dialect groups, each of which have their own recognizable subdialects; these groups include Alaskan Iñupiaq, spoken from the Seward Peninsula and north; Western Canadian Inuit (Western Canadian Inuktitut), spoken over a vast area of Central Arctic Canada from MacKenzie Coast to Repulse Bay; Eastern Canadian Inuktitut, spoken in Baffin Island, Arctic Quebec, and Labrador; and Greenlandic Kalaallisut (Greenlandic Inuktitut), spoken in Greenland (there is also a sizable population of speakers in Denmark). Characteristic of Inuit are

- lack of intonational stress as compared with Yupik;
- loss of the fourth vowel, with various important phonological traces;
- various degrees of consonant and vowel cluster simplifications (cf. Iñupiaq *aglaun* ‘pen,’ Greenlandic *allaat* ‘pen,’ in which /gl/ became /ʈ/ and /au/ became /aa/);
- a tendency to merge parts of the mood system most important in narration, with most extensive merging in Alaska and least in Greenland.

All Inuit dialects have borrowed extensively from the respective languages of colonization. Most loans are from English in Alaska and Canada (although there are also French and German loans in eastern Canada, through the influence of missionaries) and from Danish in Greenland. The status and viability of the Inuit language is quite different in the different regions. In Alaska and western Canada, the language is severely endangered, with only about 3000 speakers, almost none of whom are children. In eastern Canada, there are about 20 000 speakers, but there is widespread bilingualism in almost all age groups and a growing tendency for English to replace Inuktitut. Active efforts are under way to reverse this process, including the encouragement of Inuktitut programs in schools. In Greenland, however, over 95% of the native population of some 50 000 are speakers of Kalaallisut, and the language is thriving.

Bibliography

Bergsland K (1951). ‘Aleut demonstratives and the Aleut–Eskimo relationship.’ *International Journal of American Linguistics* 17, 167–179.

- Bergsland K (1986). ‘Comparative Eskimo–Aleut phonology and lexicon.’ *Journal de la Société Finno-Ougrienne* 80, 63–137.
- Bergsland K (1989). ‘Comparative aspects of Aleut syntax.’ *Journal de la Société Finno-Ougrienne* 82, 7–80.
- Bergsland K (1997). *Aleut grammar / Unangam Tunuganaan achixaasix*. Fairbanks: Alaska Native Language Center.
- de Reuse W J (1994). *Siberian Yupik Eskimo: the language and its contacts with Chukchi*. Salt Lake City: University of Utah Press.
- Dorais L-J (1996). *La parole Inuit: langue, culture et société dans l’arctique nord-américain*. Paris: Peeters.
- Fortescue M (1984). *West Greenlandic*. London: Croom Helm.
- Fortescue M (1998). *Language relations across Bering Strait: reappraising the archaeological and linguistic evidence*. London: Cassell.
- Fortescue M, Kaplan L & Jacobson S (1994). *Comparative Eskimo dictionary*. Fairbanks: Alaska Native language Center.
- Jacobson S (1995). *A practical grammar of the Central Alaskan Yup’ik Eskimo language*. Fairbanks: Alaska Native Language Center.
- Kleinschmidt S (1851). *Grammatik der grönländischen Sprache*. Berlin: G. Reimer.
- Krauss M (ed.) (1985). *Yupik Eskimo prosodic systems: descriptive and comparative studies*. Fairbanks: Alaska Native Language Center.
- Laughlin W S (1980). *Aleuts: survivors of the Bering land bridge*. Fort Worth, TX: Harcourt Brace College.
- Lowe R (1985). *Uummarmiut uqalungiha ilihaurrutikran-git / Basic Uummarmiut Eskimo grammar*. Inuvik, NT: COPE.
- MacLean E A (1986). *North Slope Iñupiaq grammar: first year* (3rd edn.). Fairbanks: Alaska Native Language Center.
- Marsh G & Swadesh M (1951). ‘Eskimo–Aleut correspondences.’ *International Journal of American Linguistics* 17, 209–216.
- Menovshchikov G A (1964). *Iazyk sireniskikh Eskimosov: fonetika, ocherk morfologii, teksty i slovar*. Moscow: Izdatelstvo Nauk.
- Swadesh M (1962). ‘Linguistic relations across Bering Strait.’ *American Anthropologist* 64, 1262–1291.
- Thalbitzer W (1916). *Et manuskript af Rasmus Rask. Oversigt over det Kgl. Danske Videnskabernes Selskabs Forhandling 3*, 211–249. Copenhagen.
- Tersis N & Therrien M (eds.) (2000). *Les langues eskaléoutes: Sibérie, Alaska, Canada, Groënland*. Paris: CNRS.
- Uhlenbeck C C (1935). ‘Eskimo en Oer-Indogermaansch.’ *Mededeelingen der Koninklijke Akadademie van Wetenschappen Afdeling Letterkunde*, Deel 77, Ser. A.6, 179–196.
- Veniaminov I (1846). *Opyt grammatiki aleutsko-lis’evskago iazyka*. St Petersburg: Akademiia Nauk.

Esperanto

J C Wells, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Esperanto is a constructed language intended for international use. Originating as an artificial language, it is unique in that it has enjoyed sufficient success to have acquired a speech community and even to have undergone a degree of creolization. Consequently, the epithet ‘artificial’ is arguably no longer applicable. Unlike computer languages and codes, Esperanto generally satisfies the criteria for recognition as a form of natural language. However, its proponents’ hopes of its being generally adopted for international use have not been realized, and its future must be seen as uncertain.

Background

The creator of Esperanto was Ludovic Lazar Zamenhof, a Jewish oculist of Warsaw, who used the pseudonym Doktoro Esperanto, ‘the one who hopes’. His *Lingvo internacia* was first published in 1887, in Russian. By the beginning of the 20th century, Esperanto, as the language itself quickly came to be known, had been taken up in France, Germany, and elsewhere; the London Esperanto Club (still in existence) was founded in 1903. The participants at the first International Esperanto Congress, held in 1905 in Boulogne-sur-Mer, proclaimed the essential linguistic basis (*Fundamento*) of the language as inviolable. Support has subsequently spread to most parts of the world, including Japan, China, and Brazil (although the movement remains very weak in most of the Third World). Both Stalin and Hitler saw the internationalist and idealist values of Esperanto as a threat and launched persecutions of its supporters (Lins, 1988). As of 2004, the number of speakers of Esperanto is unknown, but variously estimated as between one or two hundred thousand and several million. Universala Esperanto-Asocio, with headquarters in Rotterdam, has members in over 100 countries. Esperanto speakers in the news recently include 1994 Nobel laureate in economics Reinhard Selten, 1996 World Chess Champion Zsuzsa Polgar, and Tivadar Soros, father of financier George Soros.

The annual World Esperanto Congress – held entirely in Esperanto – regularly attracts participants in the thousands.

It must be emphasized that Esperanto is a real language, both spoken and written, successfully used as a means of communication between people who have no other common language. For the great majority of its users, of course, it is a second

language, learned well after the acquisition of the L1, so that levels of competence vary widely. However, for some speakers, children of parents who use Esperanto as a family language, it is a native language or mother tongue, normally in a bilingual or trilingual relationship with the language of the local community or other parental language(s). There is no other case in linguistic history of something that started as an intellectual scheme, a project on paper, being transformed into a language with native speakers of the second and, indeed, the third generation.

The traditional aim of the Esperanto movement is the adoption of Esperanto as L2 for all mankind. The chief arguments for this can be summarized by saying that Esperanto is easy to learn and politically neutral.

- *Ease of learning* is a consequence of the complete regularity of the language: grammatical rules have no exceptions, and the agglutinative morphological structure (discussed in the morphology section) makes vocabulary acquisition much faster than for other languages. As a result, it is claimed, Esperanto can be learned three to ten times as fast as national or ethnic languages (although obviously the rate of progress depends, as always, on many factors, including the learner’s L1).
- *Political neutrality*: Esperanto belongs to no particular nation or ethnic group. This, it is claimed, makes it politically a better choice for an international common language than English (seen by many as irredeemably attached to parochial U.K. or U.S. values) or other national languages.

Opposition to Esperanto is often more emotional than rational. Serious critics, however, argue that Esperanto is a language without culture (although supporters of Esperanto would dispute this, pointing to over a hundred years’ literary activity, including a substantial body of original poetry), and second, that it is too European (though all alternative solutions to the question of an international language are even more so). In any case, it is claimed, the economic, social, and political pressures in favor of the choice of English for international use are by now overwhelming.

In light of the perceived success of English in filling the role for which Esperanto was intended, one group (not the majority) within the Esperanto movement has redefined its aims as securing linguistic rights as a ‘stateless diaspora linguistic collective.’ It sees the speakers of Esperanto as being comparable to the speakers of other endangered or minority languages.

Pronunciation and Orthography

The phoneme system comprises 23 consonants and 5 vowels. The consonants are plosives /p b t d k g/, fricatives /f v s z ʃ ʒ x h/, affricates /ts tʃ dʒ/, nasals /m n/, liquids /l r/ and glides /j w/ (the latter only after a vowel). The vowels are /i, e, a, o, u/. Word stress is always penultimate. Orthography is strictly phonemic, using the Latin alphabet and including the following letters bearing diacritics: *ĉ*/tʃ/, *ĝ*/dʒ/, *ĥ*/x/, *ĵ*/ʒ/, *ŝ*/ʃ/, and *ŭ*/w/.

An international standard of 'good' pronunciation has by now evolved and includes the avoidance of marked interference from speakers' L1s. Intonation follows general, mainly European, models without parochialisms.

Morphology and Syntax

Esperanto syntax and morphology show strong Slavic influences. Its morphemes are invariant and can be almost indefinitely recombined into different words. The internal word structure has affinity with agglutinative languages in that all words (other than function words) consist of a stem plus a grammatical ending (-o for nouns, -a for adjectives, -i for infinitive verbs, etc.: *telefon-o* 'a telephone', *telefon-a* 'telephonic', *telefon-i* 'to telephone'). The plural ending is -j (*telefonoj* 'telephones'). There is an accusative case -n, used for the direct object. Adjectives agree with nouns in number and case: *mi vidis grandan hundon* 'I saw a big dog,' *du grandaj hundoj atakis min* 'two big dogs attacked me.' There are three verb tenses, -as present, -is past, and -os future, plus -us conditional and -u interjective/volitive. The definite article, invariant, is *la*; there is no indefinite article. The grammar is entirely regular and without exceptions.

The normal word order is SVO. Determiners and adjectives usually precede nouns, and the language is prepositional. However, the morphological marking of the accusative means that the order of constituents is flexible and can be rather freely scrambled. The last sentence quoted has emphatic or stylistic variants such as *atakis min du hundoj grandaj*.

The stem of a word may be a single root (base morpheme) or a combination of roots and/or affixes: for example,

- (1) *parol-ant-o*
speak-pres.ppl-NOUN
'speaker'
- (2) *parol-em-a*
speak-tending-ADJ
'talkative'

- (3) *mar-bord-o*
sea-edge-NOUN
'seaside'

Each morpheme (root, affix, ending) is in principle unvarying in form and meaning; the meaning of a compound of affixed form is the sum of the meanings of its constituent elements. A 20th-century development is the increasing use of affixes as independent stems, for example, *em-o* 'inclination.'

Vocabulary

The lexical material was chosen by Zamenhof to be as international as possible. Some three-fourths of the basic roots are of Romance origin, with the remainder mostly Germanic or Slavic; often they are common to several source families (e.g., *dom-* 'house'). Classical and international roots are readily incorporated, though adapted to Esperanto phonology and morphology; thus since *teatr-o* is 'theater,' 'theatrical' must be *teatr-a*. Given *telefon-o* 'telephone,' *telefone* is automatically the adverb 'by telephone' and *telefon-i* the verb 'to telephone.' The extensive use of affixes means that the vocabulary is highly structured and comparatively easy to learn. From *vidi* 'to see,' dozens of derivatives such as *videbla* 'visible,' *videbligi* 'to render visible,' and *nevidebleco* 'invisibility' can be regularly and predictably formed.

General

The language is monitored by an Academy (Akademio de Esperanto). This does not, however, inhibit individual speakers from constant linguistic creativity. Many Esperanto speakers enjoy debating linguistic issues and arguing about vocabulary innovations. It is only recently that computing terminology has settled down, with *komputilo* (regularly formed by the suffix -il- 'instrument,' cf. *hakilo* 'ax') as the term for 'computer' and technical expressions such as *elŝuti* 'to download.' Popular spoken Esperanto occasionally deviates from the written norm. For example, although the official form of the word for 'tax' is *imposto*, and this is the only form given in dictionaries, in spoken Esperanto the form *impoŝto* can be heard from time to time (because of contamination from *poŝto* 'post, mail'). The standard and written word for 'plastic' is *plasta*, but in conversation *plastika* is also heard. The existence of such 'incorrect' spoken forms can be interpreted as an indication that the language is truly socially embedded ('living').

With a speech community scattered around the world, but nevertheless commanding great feelings

of loyalty from its speakers, Esperanto is indeed comparable in some ways to diaspora languages such as Romani, Yiddish, or Armenian. In its lack of official government status, it is like creoles and many ethnic minority languages. It remains to be seen whether in the last analysis it is a linguistic curiosity doomed to disappear or a brilliant idea whose moment has not yet come.

Bibliography

- Fettes M (2002). *An update on Esperanto*. <http://www.uea.org>.
 Forster P G (1982). *The Esperanto movement*. The Hague: Mouton.
 Janton P (1993). 'Esperanto.' In Tonkin H (ed.) *Language, literature, and community*. Albany: State University of New York Press.
 Lindstedt J (1995). *Esperanto kiel denaska kaj familia lingvo* [Esperanto as a native and family language]. www.helsinki.fi.
 Lins U (1988). *La danĝera lingvo*. Gerlingen: Bleicher.
 Pokrovskij S (2003). *Komputada leksikono* [Computing lexicon]. <http://www.iis.nsk.su/>.

- Tonkin J (1977). *Esperanto and international language problems: a research bibliography*. Washington, DC: Esperantic Studies Foundation.
 Tonkin J & Fettes M (1996). 'Esperanto studies: an overview.' In *Esperanto documents* 43. Rotterdam: Universal Esperanto Association.
 Wells J C (1969). *Teach yourself concise Esperanto and English dictionary*. London: English Universities Press.
 Wells J C (1978). *Lingvistikaj aspektoj de Esperanto* [Linguistic aspects of Esperanto]. Rotterdam: Universala Esperanto-Asocio.
 Wennergren B (2003). *Plena manlibro de Esperanta gramatiko* [Complete manual of Esperanto grammar]. <http://www.bertilow.com>.

Relevant Websites

- <http://gxangalo.com>
<http://www.akademio-de-esperanto.org>
<http://www.esperanto.net>
<http://www.esperantoeducation.com/culture.html>
<http://www.esperanto-gb.org>
<http://www.esperanto-usa.org>
<http://www.phon.ucl.ac.uk/home/wells/esperanto-info>
<http://www.uea.org>

Estonian

R T Harms

© 2006 Elsevier Ltd. All rights reserved.

Estonian, a Baltic–Finnic language within the Uralic family, is represented by over a million speakers, as the official language of Estonia as well as actively maintained by émigrés in Sweden, Canada, and elsewhere. The diverse Estonian dialects may be divided into a northern group and a southern group, with Tallinn and Tartu as their respective cultural centers. Standard Estonian, in large part based on the northern dialects, underwent significant adjustment at the hands of language planners during the twentieth century to provide a compromise literary medium suitable for all areas. In contrast with nationalist movements in other countries, foreign loanwords have been tolerated, in part motivated by a desire to facilitate access to Western European culture. The resulting standardization, however, has commonly forced Estonians to consult an extensive orthographic manual for spelling and morphological norms.

The history of Estonian must be projected from reconstruction, as apart from fragments dating from ca. 1220, the first written texts appeared in the

sixteenth century, beginning with the Kullamaa prayers in the 1520s. In its development the interaction of inherited features with those of contact languages (including governing superstrata, especially German) has played a major role in shaping its unique structure.

Estonian phonology is characterized by an abundance of vowels (/i ü u e ö o ä a/ plus a midcentral unrounded /õ/), few consonants (/p t k s h m n l r y v/ plus palatalized /tʲ sʲ nʲ lʲ/—and /f š/ in loans), and a richness of stressed syllable structures resulting from contrastive vowel length (*a* versus long *aa*) tautosyllabic vowel clusters (e.g., *ea*, *eo*, *õa*), and geminate consonants (*ala* versus *alla*). The writing system is based on the Latin alphabet, augmented by diacritics (*ü*, *õ*), but palatalization is not represented. Initial stress is the rule, but loans often preserve noninitial stress (*hotell*), and multiple word stresses may result from morphology (*tulemata*) < /tule + ma + tta/—with a second stress, often stronger, on the third syllable).

Typologically, the most salient feature of Estonian is its extrasegmental quantitative prosody, by which stressed heavy syllables contrast in two types of quantity and tonal contour; for example, segmental /saata/

saada together with prosodic quantity may yield [sa:ta] 'send!' (long *a*) or [sa::ta] 'to become' (overlong *a*); *linna*, [linna] 'city (gen.)' or [lin:na] 'into city'; *osta*, [osta] 'buy!' or [os:ta] 'to buy' (contrast here /sata/ *sada* 'hundred,' *lina* 'linen,' with light first syllables). The first syllable of a form such as /saatta/ *saata* 'to send' may occur with the overlong prosodic quantity without requiring individual component segments (/aa/ or /tt/) to be overlong. Prosodic quantity is noted in spelling only for intervocalic *p*, *t*, *k*; e.g., [ata] as *ada*, [atta] as *ata*, and [at:ta] as *atta*.

The palatalized coronal consonants /tʸ sʸ nʸ lʸ/ are of special interest, since these reflect an earlier Umlaut-like process (contrasting typologically with

their Slavic counterparts in the absence of a prominent offset transition). This palatalization produces a *y*-like onset transition from a preceding stressed vowel (triggered by an [earlier] *i* in the following syllable); e.g., /kasʸk/ *kask* = [kaʸsk] < **kaski* 'birch' (contrast the absence of palatalization in *kaks* < **kaksi* 'two').

Bibliography

- Kurman G (1968). *The development of written Estonian*. Bloomington: Indiana University Press.
 Raun A & Saareste A (1965). *Introduction to Estonian linguistics*. Wiesbaden: Harrassowitz.

Ethiopia as a Linguistic Area

J Crass, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Ethiopian linguistic area is probably the most famous linguistic area in Africa. It is often the only African linguistic area discussed to any extent in general works dealing with language contact and areal linguistics (Masica, 1976; Thomason and Kaufman, 1988; Thomason, 2001). Most scholars refer to this area as the Ethiopian language area (Ferguson, 1976; Sasse, 1986; Hayward, 1991; Zaborski, 1991, 2003; Tosco, 1994; Crass, 2002). This term, however, is problematic in several respects. First, the English translation of the German term *Sprachbund* is linguistic area, convergence area, or diffusion area (Campbell, 1994: 1471). Second, the area includes Eritrea, which was a province of Ethiopia until it became an independent state in 1993. Third, a certain number of features are found beyond Ethiopia and Eritrea in languages spoken in neighboring countries, such as Djibouti, Somalia, and Sudan, and even beyond. Some scholars have taken these facts into account, at least to some extent. Hayward (2000: 623) uses the term Ethio-Eritrean Sprachbund; Zaborski (2003: 62) proposes North East African Language Macro-Area. Bender (2003a: 4) argues that the terms Northeast Africa area, Horn of Africa Language Area, and Erythraean Area "all have their pluses and minuses" and stresses that "now we must modify it to Ethiopia-Eritrean Area." In the present article, the term Ethiopian linguistic area (ELA) is

used in order because (1) language area is not the correct term in areal linguistics and (2) the core of the area is Ethiopia.

Most of the approximately 80 languages spoken in Ethiopia and Eritrea belong to three language families of the Afroasiatic phylum: Semitic, Cushitic, and Omotic. A number of languages in the west and in the southwest belong to various families of the Nilo-Saharan phylum.

According to a widely accepted view, Semitic-speaking peoples arrived in the Horn of Africa at the end of the first millennium B.C. by crossing the Red Sea after having left their home on the Arabian Peninsula. They migrated into the area of today's Ethiopia and Eritrea and underwent extensive linguistic and extra-linguistic influence due to contact with Cushitic-speaking peoples (Ullendorff, 1955). A contradictory view considers Ethiopia and Eritrea to be the original homeland of Semitic-speaking people (Murtonen, 1967; Hudson, 1977). This view is based on the observation that the linguistic diversity among Semitic languages in Ethiopia and Eritrea is much greater than elsewhere.

The proposed features defining the ELA differ considerably from author to author, and the validity of these features has been discussed by the authors previously mentioned. Recently, even the existence of an ELA was rejected by Tosco (2000).

Research History

Leslau (1945) and Moreno (1948) are two early attempts to describe the influence of Cushitic languages on Ethiopian Semitic languages. The first to

claim the existence of a linguistic area in “Ethiopia and the various Somalilands” was Greenberg (1959: 24). He argues that this area is marked by “relatively complex consonantal systems, including glottalized sounds, absence of tone, word order of determined followed by determiner [*sic*], closed syllables, and some characteristic idioms.” According to Heine (1975: 41ff.), who deals with word order, Ethiopia is part of “probably the largest convergence area in Africa, stretching in a broad belt from the Lake Chad region in the west to the Red Sea and the Indian Ocean in the east.”

Ferguson (1970, 1976) was the first to describe the ELA in more detail. The second of these articles, with an extended database and improvements and corrections, is still the reference work for all scholars. Therefore, Ferguson (1976) is the starting point when in the following discussion of phonological and grammatical features. Ferguson discusses eight phonological and eighteen grammatical areal features for 18 languages, including Arabic and English. He is of the opinion that the “languages of Ethiopia constitute a linguistic area in the sense that they tend to share a number of features which, taken together, distinguish them from any other geographically defined group of languages in the world” (Ferguson, 1976: 63f.). He stresses that “some of these shared features are due to genetic relationship [...], while others result from the process of reciprocal diffusion among languages which have been in contact for many centuries” (Ferguson, 1976: 64).

Zaborski (1991: 124) criticizes Ferguson’s selection of languages and features. He is of the opinion that the languages were “rather random[ly] selected” and that “most of the alleged areal features are not really areal but of common genetic origin.” Hayward (2000: 623) argues that a number of Ferguson’s features are “characteristic of most languages of this region”; five of these features he considers to be “very widespread.” Some articles deal with only one areal feature: Appleyard (1989) discusses relative verbs in focus constructions, Tosco (1994) deals with case marking, and Tosco (1996) deals with extended verb paradigms in the Gurage-Sidamo subarea.

Tosco (2000) is a controversial paper. In it, he denies the existence of an ELA because of the genetic relatedness of Ethiopian Semitic and Cushitic languages, the unilateral diffusion from Cushitic to Ethiopian Semitic, and the occurrence of features in related languages that do not belong to the ELA. Four recent papers, Bender (2003b), Crass (2002), Crass and Bisang (2004), and Zaborski (2003: 62f), favor the existence of a linguistic area. Bender (2003b) argues against the conclusions in Tosco (2000) and tries to extend the ELA by testing a

number of Nilo-Saharan languages using a selection of Ferguson’s features. Crass (2002) discusses two phonological features in detail; in Crass and Bisang (2004), the discussion is extended to features such as word order, converbs, and ideophones verbalized by the verb ‘to say’. Zaborski (2003) presents the most extensive list, including 28 features that he considers to be valid for a macroarea that includes Ethiopia, Eritrea, Djibouti, Somalia, and parts of Sudan, Kenya, and even Tanzania and Uganda. Finally, Hayward (1991) deals with patterns of lexicalization shared by the three Ethiopian languages, Amharic (Semitic), West-Central Oromo (Cushitic), and Gamo (Gamo-Gofa-Dawro; Omotic). According to Hayward (1991: 140), these lexicalizations reinforce “the very real cultural unity of Ethiopia.” This topic is reopened in Hayward (2000).

The ELA is considered to be composed of several subareas. Leslau (1952, 1959) describes change in Ethiopian Semitic languages induced by contact with neighboring Highland East Cushitic languages. Sasse (1986) deals with the Sagan area in southwest Ethiopia, and Zaborski (1991: 125) gives a list of seven subareas that are composed of “smaller contact and interference units,” which he extends to nine by adding a Kenyan and an Tanzanian subarea (Zaborski 2003: 64).

Phonological Features

The eight phonological features listed by Ferguson (1976: 65ff.) are

- P1: /f/ replacing /p/ as the counterpart of /b/
- P2: palatalization of dental consonants as a common grammatical process in at least one major word class
- P3: the occurrence of ejectives (in Ferguson’s terminology, glottalic consonants)
- P4: the occurrence of an implosive /d/
- P5: the occurrence of pharyngeal fricatives
- P6: the occurrence of consonant germination
- P7: the occurrence of central vowels
- P8: the occurrence of an epenthetic vowel

Ferguson’s list has since been criticized by several scholars. Zaborski (1991: 114 fn. 3) considers only P3 and “with reservations” P2 to be “really areal.” In his more recent paper, Zaborski (2003: 62) lists only four phonological features for ELA. In addition to P3 and P6, he argues that “labialized consonants are frequent [and that] some palatalized consonants are innovations.” Tosco (2000: 341), in contrast, is of the opinion that P1, P2, P3, and P5 are genetically inherited within the Afroasiatic phylum; that P4, P7, and P8 are restricted

to one or two language families; and that P6 is widespread in both the Afroasiatic and the Nilo-Saharan phyla. But according to Bender (2003a: 3, 2003b: 32), P2 and P6 are typological features, P5 is too limited, and P8 “is vacuous because consonant clusters are rare” (Bender, 2003a: 3). P1, P3, P4, and P7, however, are “fairly idiosyncratic and easy to check” (Bender, 2003b: 32). Hayward (2000: 623) explicitly mentions only P6, which he considers to be areal. Crass (2002) discusses the ejectives and pharyngeal fricatives (P3 and P5) in detail. Because both features are genetically inherited within the Afroasiatic phylum, Crass argues that their occurrence (in the case of ejectives) and nonoccurrence (in the case of pharyngeal fricatives) can be considered areal features. Reconstructions of different stages of proto-languages of Afroasiatic show that ejectives were lost over the course of time (for details, see Crass, 2002: 1683); in recent times, however, ejectives were reimported into most of the languages via contact. In Proto-Highland East Cushitic, for example, only one ejective is attested, the velar ejective. In most of the modern Highland East Cushitic languages, however, four ejectives occur as phonemes: the dental, the postalveolar affricate, the velar, and to a lesser extent the labial ejective (Hudson, 1989: 11). In the Agaw languages (Central Cushitic), ejectives occur predominantly in loan words from Amharic and Tigrinya (Tigrigna) and their phonemic status is problematic (cf. Appleyard, 1984: 34).

The reasons for the nonoccurrence of pharyngeal fricatives in most of the languages of central Ethiopia are unclear. The nonoccurrence may be due to language contact or due to language-internal change. Tosco (2000: 343) supports Crass’s idea in arguing that the nonoccurrence of pharyngeal fricatives “could identify a smaller ‘central Ethiopian area’ ... in which pharyngeal consonants are either dropped or reduced.”

Grammatical Features

Ferguson’s (1976: 69ff.) grammatical features are

- G1: SOV word order
- G2: subordinate clauses precede main clauses
- G3: converb
- G4: postpositions
- G5: quotation marked by the verb ‘to say’ (in Ferguson’s terminology, quoting clauses)
- G6: “compound verbs ... consisting of a noun-like or interjection-like ‘preverb’ plus a semantically colourless auxiliary, commonly the verb ‘to say’” (Ferguson, 1976: 71f.)

- G7: negative copula
- G8: singular used with numbers
- G9: possessive suffixes identical or nearly identical with object suffixes added to the verb
- G10: masculine/feminine gender distinction in the second- and third-person singular of pronouns and verbs
- G11: identical subject prefixes for the second-person masculine singular and the third-person feminine singular marking a certain tense that contrast with subject suffixes forming other tenses
- G12: for many words, a consonantal skeleton carrying the lexical meaning and a pattern of vowels carrying the grammatical meaning
- G13: reduplication for forming intensive verbs and plurals of adjectives
- G14: plural formation by change of the pattern (e.g., ablaut, called broken plurals)
- G15: an independent and a subordinate form of the imperative
- G16: plural nouns agree with a feminine singular adjective, verb, or pronoun
- G17: the imperative of the verb ‘to come’ is formed either from “a totally different stem [...] or with an exceptional formation” (Ferguson, 1976: 74)
- G18: the unmarked form of a noun is not singular in number but plural or collective (Ferguson, 1976: 74)

Zaborski (1991: 124) considers G1 to G6 to be areal features and G7 to G18 to be due to genetic origin. Furthermore, he adds two features that he considers areal: (1) adjectives precede substantives and (2) main verbs precede auxiliaries. Hayward (2000: 623) is of the opinion that G1, G3, G6, and G15 are “very widespread.” According to Bender (2003a: 3, 2003b: 32), G2, G3, G4, and G9 “are implicational consequences of SOV order,” G13 and G18 are “too typological,” and G10 to G12, G14, and G16 are Afroasiatic, “especially Semitic idiosyncrasies.” In addition, G7 “looks like a good choice but turns out [...] to be inadequately defined.” Bender seems to consider the features G1, G5, G6, G8, G15, and G17 to be good candidates for areal features.

Hayward (1991) and Tosco (2000) correctly stress that G2 and G4 have a relation to G1 and therefore cannot count as individual features. Within this context, Campbell (1994: 1471) raises the question of the weight of “a trait so central to the grammar” when it is counted only as a single feature. Also, G3, contrary to Bender’s opinion, is not related to G1 (cf. Bisang, 2001) and is found in an area exceeding the ELA. Tosco (2000) considers G3 to have spread into Semitic languages due to Cushitic influence. According to Tosco, this holds true also for G5, G6, G8, G13, and G15. The features G11, G12, and G14 are strongly

“Semitic-biased,” and G17 and G18 are Afroasiatic features.

The huge list of areal features presented by Zaborski (2003: 62f.) contains features of differing quality. Unfortunately, in most cases Zaborski simply names the features without any discussion. A number of features relate to the basic SOV word order; examples are (1) dependent clauses precede main clauses, (2) main verbs precede auxiliary ones, (3) adjectives precede nouns that they define and (4) possessor (genitive) precedes the possessed. Other features are trivial or represent frequent grammaticalizations, such as (1) relative clauses are frequent, (2) cleft sentences are frequent, (3) subject is in the oblique case, and (4) postpositions start functioning as new case endings. However, there are a few interesting features that need further study, such as (1) quoting clauses and a lack (or at least limited use) of indirect speech and (2) a considerable number of different ‘to be’ auxiliary verbs.

Lexical Features

Hayward (1991) distinguishes three categories of lexicalizations, which he exemplifies with data from Amharic, Oromo, and Gamo. These categories are (1) single-sense lexicalizations, (2) lexicalizations with two or more distinct senses, and (3) lexicalizations involving similar derivations. The first category comprises “single-sense lexicalizations of typically indigenous concepts” (Hayward 1991: 145), the second category lexicalizations “showing inter-linguistic matching across the three languages” (Hayward 1991: 148), and the third category lexicalizations with a “similar (parallel) ‘derivational pathway’” (Hayward 1991: 150). To the first category belong mainly nouns, such as lexical items for seasons of the year, categories of terrain, categories of dung/excrement, supercategories for birds, types of borrowing, and the skin-color classification of people of the region. Furthermore, this category includes the suppletive imperative of the verb ‘to come’ (Ferguson’s feature G17) and particles with the meaning ‘take this!’, which have no obvious etymological relationship to a verb. The second category, lexicalizations with two or more distinct senses, predominantly comprises verbs and some nouns; examples are the verb that has the basic meaning ‘hold, catch’ and the secondary meaning ‘start, begin’ and the verb that has the basic meaning ‘play’ and the secondary meaning ‘chat’. The third category includes verbal derivations, compound verbs (Ferguson’s feature G6), possessive constructions including two NPs, and idiomatic expressions. Examples of verbal derivations are the causative of the verb ‘want’, which has

the meaning ‘need’; the causative of the verb ‘enter’, which has the meaning ‘marry’; and the causative of the verb ‘pass the night’, which has the meaning ‘administer’. Examples of compound verbs are ‘become silent’, ‘hurry up’, and ‘jump up suddenly’. Possessive constructions that include two NPs have a word-by-word meaning and a metaphorical meaning. Examples are ‘son of man/people’, with the metaphorical meaning ‘mankind, human being’, and ‘land of man/people’, with the metaphorical meaning ‘foreign country’. Examples of idiomatic expressions are ‘regain/recover control, take courage’, which is composed of the noun ‘heart’ and the verb ‘return (INTRANS)’, and ‘catch cold’, in which the noun ‘cold’ is the subject and the experiencer is the object of the verb ‘catch’.

Summary

The ELA is characterized by phonological, grammatical, and lexical features. Their areal status is not accepted by all scholars; furthermore, the existence of the ELA itself is not generally accepted. Most of the features are found in languages spoken in the highlands of Ethiopia. The more distant a given language is from this core area, the fewer features it has. However, because only a relatively small number of languages have been adequately described, these findings must be considered preliminary.

Bibliography

- Appleyard D (1984). ‘The internal classification of the Agaw languages: A comparative and historical phonology.’ In Bynon J (ed.) *Current progress in Afro-Asiatic linguistics: Papers of the Third International Hamito-Semitic Congress*. Philadelphia: John Benjamins. 33–67.
- Appleyard D (1989). ‘The relative verb in focus constructions: An Ethiopian areal feature.’ *Journal of Semitic Studies* 34.2, 291–305.
- Bender M L (2003a). Ethiopian language area. Unpublished manuscript. 1–5.
- Bender M L (2003b). Northeast Africa: A case study in genetic and areal linguistics. *Annual Publication in African Linguistics* 1, 21–45.
- Bisang W (2001). ‘Finite vs. non finite languages.’ In Haspelmath M, König E, Oesterreicher W & Raible W (eds.) *Language typology and language universals*. Berlin: De Gruyter. 1400–1413.
- Campbell L (1994). ‘Grammar: Typological and areal issues.’ In Asher R E & Simpson J M Y (eds.) *The encyclopedia of language and linguistics*, vol. 3. New York: Pergamon. 1471–1474.
- Crass J (2002). ‘Ejectives and pharyngeal fricatives: Two features of the Ethiopian language area.’ In Baye Y et al. (eds.) *Ethiopian studies at the end of the second millennium. Proceedings of the 14th International*

- Conference of Ethiopian Studies*. Addis Ababa: Institute of Ethiopian Studies. 1679–1691.
- Crass J & Bisang W (2004). 'Einige Bemerkungen zum äthiopischen Sprachbund und ihre Relevanz für die Areallinguistik.' In Bisang W, Bierschenk T, Kreikenbom D & Verhoeven U (eds.) *Kultur, Sprache, Kontakt*. Würzburg: Ergon. 169–199.
- Ferguson C A (1970). 'The Ethiopian language area.' *Journal of Ethiopian Studies* 8.2, 67–80.
- Ferguson C A (1976). 'The Ethiopian language area.' In Bender M L, Bowen J D, Cooper R L & Ferguson C A (eds.) *Language in Ethiopia*. London: Oxford University Press. 63–76.
- Greenberg J (1959). 'Africa as a linguistic area.' In Bascom W R & Herskovits M J (eds.) *Continuity and change in African languages*. Chicago: University of Chicago. 15–27.
- Hayward R J (1991). 'À propos patterns of lexicalization in the Ethiopian language area.' In Mendel D & Claudi U (eds.) *Ägypten im afro-asiatischen Kontext. Aufsätze zur Archäologie, Geschichte und Sprache eines unbegrenzten Raumes. Gedenkschrift Peter Behrens*. Köln: Institut für Afrikanistik. 139–156.
- Hayward R J (2000). 'Is there a metric for convergence?' In Renfrew C, McMahon A & Trask L (eds.) *Time depth in historical linguistics*, vol. 1. Cambridge, UK: McDonald Institute for Archeological Research. 621–640.
- Heine B (1975). 'Language typology and convergence areas in Africa.' *Linguistics* 14.4, 27–47.
- Hudson G (1977). 'Language classification and the Semitic prehistory of Ethiopia.' *Folia Orientalia* 18, 119–166.
- Hudson G (1989). *Highland East Cushitic dictionary*. Hamburg: Buske.
- Leslau W (1945). 'The influence of Cushitic on the Semitic languages of Ethiopia: A problem of substratum.' *Word* 1.1, 59–82.
- Leslau W (1952). 'The influence of Sidamo on the Ethiopic languages of Gurage.' *Language* 28.1, 63–81.
- Leslau W (1959). 'Sidamo features in the South Ethiopic phonology.' *Journal of the African and Oriental Society* 79.1, 1–7.
- Masica C P (1976). *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.
- Moreno M M (1948). 'L'azione del cuscitico sul sistema morfologico delle lingue semitiche dell'Etiopia.' *Rassegna di Studi Etiopici* 7, 121–130.
- Murtonen A (1967). *Early Semitic: A diachronical inquiry into the relationship of Ethiopic to other so-called South-East Semitic languages*. Leiden: Brill.
- Sasse H-J (1986). 'A southwest Ethiopian language area and its cultural background.' In Fishman J A, Tabouret-Keller A, Clyne M, Krishnamurti B & Abdulaziz M (eds.) *The Fergusonian impact: In honour of Charles A. Ferguson on the occasion of his 65th birthday*, vol. 1. Berlin: Mouton de Gruyter. 327–342.
- Thomason S G (2001). *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason S G & Kaufman T (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tosco M (1994). 'On case marking in the Ethiopian language area (with special reference to subject marking in East Cushitic).' In Brugnatelli V (ed.) *Sem, Cam, lafet*. Milano: Centro Studi Camito-Semitici. 225–244.
- Tosco M (1996). 'The northern Highland East Cushitic verb in an areal perspective.' In Griefenow-Mewis C & Voigt R M (eds.) *Cushitic and Omotic languages. Proceedings of the third international symposium*. Köln: Köppe. 71–99.
- Tosco M (2000). 'Is there an "Ethiopian language area"?' *Anthropological Linguistics* 42.3, 329–365.
- Ullendorff E (1955). *The Semitic languages of Ethiopia: A comparative phonology*. London: Taylor's.
- Zaborski A (1991). 'Ethiopian language subareas.' In Pilażewicz, Stanisław & Rzewuski E (eds.) *Unwritten testimonies of the African past. Proceedings of the International Symposium*. (Orientalia Varsoviensia 2) Warsaw: Wydawnictwa Uniwersytetu Warszawskiego. 123–134.
- Zaborski A (2003). 'Ethiopian language macroarea.' *Sprawozdania z Posiedzeń Komisji Naukowych* 45.2, 60–64.

Ethiopian Semitic Languages

A Wetter, Johannes Gutenberg University,
Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

The Ethiopian Semitic (ES) languages form a subdivision of the Semitic language family that is a branch of the Afroasiatic superfamily. This subfamily comprises about 23 languages. Two languages (Tigre and Dahlik) are spoken only in Eritrea. Tigrinya (Tigrigna) is spoken in Eritrea and Ethiopia and all the others are in Ethiopia. (Unless otherwise

specified, figures of speakers cited in this article are taken from the Ethnologue website.)

Classification and Demography

The still widely accepted classification of ES is based on shared "innovations due to internal reasons" (Hetzron, 1972: 13) and has two main divisions, a northern (NES) and a southern branch (SES). The northern group is formed by only four languages: the ancient Ge'ez (now extinct), Tigrinya (1 900 000 in Eritrea and 3 225 000 in Ethiopia), Tigre (800 000),

and the recently identified Dahlik (1000) by Simeon-Senelle (2000). The southern branch, in contrast, comprises most of the languages and has a more complex composition, with several subgroups. This branch, according to Hetzron (1972), is classified into Transversal and Outer South ES. The languages forming the Transversal South ES are spoken in the central and center-eastern parts of Ethiopia and is subdivided into Central Transversal, consisting of Amharic (spoken as a first and second language by 80% of the 70 million total population of Ethiopia; cf. Meyer and Richter, 2003) and Argobba (10 000), and Eastern Transversal, consisting of the Eastern Gurage languages, i.e., Silt'e, Zay, and Wolane (together 900 000) and Harari (22 000), spoken in the city of Harar. The second subgroup, Outer South ES, contains all other languages that are divided into a 'n-group,' which consists of Gafat, a non-Gurage language that is extinct, and the North Gurage languages that comprise Goggot, Soddo (Gurage, Soddo), and, according to Leslau (1969) and Demeke (2001), Mäsqaṇ, (together around 300 000) and a 'tt-group,' which comprises about nine different languages (Muher, Ezha, Chaha, Gumär, Gura, Gyeto, Ennemor, Endegen, and Endär, together 800 000). These languages are spoken in a comparatively small area south of Addis Ababa. The term Gurage therefore refers to a group of peoples with a common cultural and historical background who speak different South Ethiosemitic languages that do not necessarily constitute a single genetic group.

The geographical origin of the ES languages is controversial. The traditional hypothesis proposes an origin in the Near East: Semitic-speaking settlers from Southern Arabia crossed the Red Sea and introduced the ancestor of modern ES languages in Northeast Africa around 1000 B.C. (cf. Ullendorff, 1955). In contrast, there is a more convincing hypothesis that sets the emergence of Semitic languages in the context of the Afroasiatic language family and proposes an origin in Ethiopia proper (cf. Murtonen, 1967; Drewes, 1958; Hudson, 1977, among others).

Phonology

Ejective consonants (p', t', s', c', and k') are widespread in all ES languages and seem to represent an earlier stage of Semitic emphatic consonants. Pharyngeal consonants are retained in NES, Argobba, and Harari but are lost in the remaining languages. Labialization, especially of velars, is common. Quite complex palatalization and labialization processes appear in several Gurage languages. Vowel length is attested only in Ge'ez and Eastgurage. The central

vowel *i* acts as an epenthetic vowel in many SES languages. Gemination of consonants is distinctive.

Morphology

The verbs are categorized in different types depending on the gemination of the penultimate consonant and the vowel quality in different conjugational paradigms. The different behavior of gemination in type A verbs is one of the distinguishing features between NES and SES languages. The former geminates in the perfective whereas the latter does so in the imperfective.

Verbal negation is marked by prefixed morphemes, which may differ for different aspects. In addition to the two main aspect forms, perfective and imperfective, there are a number of compound tense constructions formed with auxiliaries (built with the verb of existence *√blw*). A special form of verbs are the so-called compound verbs. Their composition has the form ideophone + conjugated form of the verb 'to say'; the first carrying the semantic content and the latter the grammatical functions such as person or aspect.

The use of converbs for verbal subordination (expression of sequence or adverbial concept) is common. They appear generally in two different forms: In NES languages and Amharic/Argobba gerundive forms are used, whereas in the remaining SES languages there are a number of different suffixes attached to different verb forms such as perfective and imperfective.

The relative is marked with a morpheme that is prefixed to a verb that by itself precedes the head noun of the relative clause. Verbal derivation is achieved by prefixes and internal extension of the stem. Derivational prefixes are a- and as- for causative and t- for passive or middle voice.

Nominal morphology is characterized by the broken plural in NES languages and the use of suffixed plural markers in SES languages. In some languages this plural suffix marks the transnumeralis. An article derived historically from possessive marker can be observed in some languages. Marked cases are accusative, which is marked either by a pre- or by a suffix, and genitive, which is marked by position or with a prefix at the possessor that precedes the possessed.

Syntax

The most striking feature of ES language syntax is the word order, SOV, which stands in an obvious contrast to VSO in the Asian Semitic languages. This change of word order leads to a somehow

mixed typological situation, i.e., the verb is in the sentence-final position and the subordinated clause precedes the main predication.

- (1) legan azzennewga met'illew (Argobba)
tomorrow NEG.it-rained.if I-come
'If it doesn't rain I will come'

In addition to postpositions, which are typical for SOV, prepositions are widely used. Most of them are also used as conjunctions that are usually cliticized to the following constituent (t- and b- are more frequent; k-, s-, and l- are seldom used). The suffixed pronominal features for the direct object often have the form of affix, unlike those in Asian Semitic languages, which are clitics. The present tense copula is formed in many languages by the root *ne-* plus a person-marking suffix. Tigre and Ge'ez express present tense equative clauses without a copula or with a pronoun copula, as in most Asian Semitic languages (Crass *et al.*, 2004). The past tense copula in most ES languages is *neb(b)er*. *Wh*-items are *in situ* types and question words may be used in yes–no interrogatives and usually have postverbal position. Relative forms of verbs are frequently used in nominal function.

On the discursive level, cleft sentences play a major role, especially for focusing of certain parts of a sentence.

- (2a) telantena new sewiyyew yemett'aw (Amharic)
yesterday COP man-the REL.he-
came
'It was yesterday that the man came'
- (2b) sewiyyew new telantena yemett'aw
man-the COP yesterday REL.he-came
'It was the man that came yesterday'

Most of the morpho-syntactic differences between Ethiosemitic and Asian Semitic languages are assumed to be the result of language contact on the Ethiopian side, but there are not enough data to support this claim. Documentary works still need to be conducted on many of these languages.

Bibliography

- Bender M L & Hetzron R (1976). *Language in Ethiopia*. London: Oxford University Press.
- Crass J, Demeke G A, Meyer R & Wetter A (2005). 'Copular construction in Ethiopian languages.' University of Leipzig Papers on Africa (to appear).
- Demeke G A (2001). 'The Ethiosemitic languages: Reexamining the classification.' *Journal of Ethiopian Studies* 34(2), 57–93.
- Demeke G A (2003). *The syntax of Ethiosemitic*. Ph.D. diss., University of Tromsø.
- Drewes A J (1958). 'The origin of the Semitic languages of Ethiopia.' *Ethiopian Observer* 2/3, 113–115.
- Hetzron R (1972). *Ethiopian Semitic. Studies in classification*. Manchester: University Press.
- Hetzron R (1977). *The Gunnän-Gurage languages*. Naples: Istituto Orientale di Napoli.
- Hetzron R (ed.) (1997). *The Semitic languages*. London: Routledge.
- Hudson G (1977). 'Language classification and the Semitic prehistory of Ethiopia.' *Folia Orientalia* 18, 119–166.
- Hudson G (1978). 'Geolinguistic evidence for Ethiopian Semitic prehistory.' *Abbay* 9, 71–85.
- Leslau W (1969). 'Toward a classification of the Gurage dialects.' *Journal of Ethiopian Studies* 14, 96–109.
- Meyer R & Richter R (2003). *Language use in Ethiopia from a network perspective. Results from a sociolinguistic survey conducted among high school students*. Frankfurt: Peter Lang Verlag.
- Murtonen A (1967). *Early Semitic. A diachronical inquiry into the relationship of Ethiopic to the other so-called South-Semitic languages*. Leiden: Brill.
- Rose S (1997). *Theoretical issues in comparative Ethio-Semitic phonology and morphology*. Ph.D. diss., McGill University Montréal.
- Simeon-Senelle M-C (2000). 'Situation linguistique dans le sud de l'Erythrée.' In Wolff H E & Gensler O (eds.) *Proceedings 2nd World Congress of African Linguistics, Leipzig 1997*. Köln: Rüdiger Köppe Verlag. 261–276.
- Ullendorff E (1955). *The Semitic languages of Ethiopia: a comparative phonology*. London: Taylor's.

Ethnologue

G F Simons and R G Gordon Jr, SIL International, Dallas, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

Ethnologue: Languages of the World is a reference work cataloging all known languages of the present-day world. Now in its 15th edition (2005), the

Ethnologue identifies 6912 living languages, both spoken and signed. These are distinct languages that have living mother tongue speakers. A few hundred recently extinct languages are documented as well.

For over 50 years, the *Ethnologue* has been compiled and published by SIL International, a non-profit, nongovernmental organization that studies, documents, and assists in developing the world's

lesser-known languages. Information comes from a variety of sources including reliable published sources, a network of field correspondents, and numerous personal communications that are confirmed by consulting published sources or the network of correspondents. The editorial staff processes approximately 10 000 updates to the database every year.

History of the *Ethnologue*

The *Ethnologue* was founded by Richard S. Pittman, who was motivated by the desire to share information on language development needs around the world with his colleagues in SIL International as well as with other language researchers. The first edition in 1951 was 10 mimeographed pages and included information on 46 languages or groups of languages. Maps were first included in the 4th edition (1953). The publication transitioned from mimeographed pages to a printed book in the 5th edition (1958). Dr. Pittman continued to expand his research through the 7th edition (1969), which listed 4493 languages.

In 1971, Barbara F. Grimes became editor. She had assisted with the *Ethnologue* since 1953 (4th edition) and took on the role of research editor in 1967 for the 7th edition (1969). She continued as editor through the 14th edition (2000). In 1971, information was expanded from primarily minority languages to encompass all known living languages of the world. Between 1967 and 1973, Ms Grimes completed an in-depth revision of the information on Africa, the Americas, the Pacific, and a few countries of Asia. During her years as editor, the number of identified languages grew from 4493 to 6809, and the information recorded on each expanded so that the published work more than tripled in size.

The 15th edition (2005) was edited by Raymond G. Gordon, Jr. It reflects an increase of 103 languages over the previous edition. Most of these are not newly discovered languages, but are ones that had been previously considered dialects of another language.

The Problem of Language Identification

Given the nature of language and the various perspectives brought to its study, it is not surprising that a number of issues prove controversial. Of preeminence in this regard is the definition of 'language' itself. Since languages do not have self-identifying features, what actually constitutes a language must be operationally defined. That is, the definition of language one chooses depends on the purpose one has in identifying a language. Some base their definition on purely linguistic grounds. Others recognize that

social, cultural, or political factors must also be taken into account.

Every language is characterized by variation within the speech community that uses it. The resulting speech varieties are more or less divergent from one another. These divergent varieties are often referred to as dialects. They may be distinct enough to be considered separate languages or sufficiently similar as to be considered merely characteristic of a particular geographic region or social grouping within the speech community. Scholars do not all share the same criteria for what constitutes a 'language' and what features define a 'dialect.' The *Ethnologue* applies the following basic criteria:

- Two related varieties are normally considered varieties of the same language if speakers of each variety have inherent understanding of the other variety at a functional level (that is, can understand based on knowledge of their own variety without needing to learn the other variety).
- Where spoken intelligibility between two varieties is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central variety that both understand can be a strong indicator that they should nevertheless be considered varieties of the same language.
- Even where there is enough intelligibility between varieties to enable communication, the existence of well-established distinct ethnolinguistic identities can be a strong indicator that they should nevertheless be considered to be different languages.

Increasingly, scholars are recognizing that languages are not always easily treated as discrete isolatable units with clearly defined boundaries between them. Rather, languages are more often continua of features that extend across both geographic and social space. The *Ethnologue* approach to listing and counting languages as though they were discrete, countable units does not mean to preclude a more dynamic understanding of the linguistic makeup of the countries of the world. In fact, particular language entries in the *Ethnologue* list known dialects and often comment on the similarity and intelligibility relationships among them. In the final analysis, however, the *Ethnologue* lists and counts languages as distinguished by the criteria named above because it serves as a baseline for those who are developing language policy and making plans for language development. It is also foundational for those, such as librarians and archivists, who would classify written and spoken materials with respect to the languages they are in, or would organize pieces of language-related information with respect to the languages they are about.

Three-Letter Language Identifiers

A distinctive feature of the *Ethnologue* over the years has been its use of three-letter codes to uniquely identify the languages of the world. Any enterprise that would categorize language-related resources so that others might effectively retrieve those resources depends on the uniform identification of the languages to which they pertain. Simply using language names for this purpose is not adequate since the same language is typically known by many names and those names change over time. Furthermore, different languages may be known by the same name. Thus, the most effective approach is to use standardized language identifiers.

Standardized language identifiers were introduced into the *Ethnologue* in 1971 by then consulting editor, Joseph E. Grimes, when he transformed the typesetting tapes for the 7th edition (1969) into a computerized database on languages of the world. The work was done at the University of Oklahoma under a grant from the National Science Foundation. In 1974, the database was moved to a computer at Cornell University where Dr Grimes was professor of linguistics; it was moved to a personal computer in 1979. Since 2000, it has been housed at the headquarters of SIL International in Dallas, Texas.

One feature of the database since its inception has been a system of three-letter language identifiers. Grimes explained this feature as follows in the 1974 final report for the grant: "Each language is given a three-letter code on the order of international airport codes. This aids in equating languages across national boundaries, where the same language may be called by different names, and in distinguishing different languages called by the same name." While the codes were used behind the scenes in the database that generated the 8th and 9th editions, it was not until the 10th edition (1984) that they appeared in the publication itself.

The 15th edition (2005) marked an important milestone in the development of the language identifiers, namely, their emergence as a draft international standard. In 1998, the International Organization for Standardization adopted ISO 639-2, its standard for three-letter language identifiers. That was based on a convergence of ISO 639-1 (its earlier standard for two-letter language identifiers adopted in 1988) and of ANSI Z39.53 (also known as the MARC language codes, a set of three-letter identifiers developed within the library community and adopted as an American National Standard in 1987). The current standard, ISO 639-2, has proven insufficient for many purposes since it has identifiers for fewer than 400 individual languages. Thus, in 2002, ISO TC37/SC2 invited SIL

International to participate in the development of a new standard based on the language identifiers in the *Ethnologue*. The new standard was to be a superset of ISO 639-2 that would provide identifiers for all known languages. Consequently, hundreds of the *Ethnologue* language identifiers were changed in order to achieve alignment with ISO 639-2. In 2004, the proposed new standard, ISO 639-3, passed the first round of balloting by national standards bodies to attain the status of Draft International Standard. The three-letter language identifiers in the 15th edition of the *Ethnologue* are thus the codes of ISO/DIS 639-3.

Endangered Languages

Language endangerment is a serious concern to which linguists and language planners have turned their attention in the last decade. For a variety of reasons, speakers of some languages are motivated to stop using their language and to use another. Parents may begin to use only that second language with their children. Eventually there may be no speakers who use the language as their first or primary language and frequently the language ceases to be used altogether and the language becomes extinct – existing, perhaps, only in recordings or written records and transcriptions. The concern about language endangerment is centered, first and foremost, around the factors that motivate speakers to abandon their language and the consequences of language death for the community of (former) speakers of that language. Since language is closely linked to culture, loss of language almost always is accompanied by social and cultural disruptions as well. Secondly, those concerned about language endangerment recognize the implications of the loss of linguistic diversity, both for the linguistic and social environment generally and for the academic community, which is devoted to the study of language more specifically.

There are two dimensions to the evaluation and characterization of endangerment: the number of speakers of the language and the number and nature of the domains in which the language is used. A language may be endangered because there are fewer and fewer people who speak that language. It may also, or alternatively, be endangered because it is being used for fewer and fewer functions. The *Ethnologue* attempts to provide data on both of these dimensions whenever it is available.

Language endangerment is a matter of degree. At one end of the scale are languages that are vigorous and perhaps are even expanding in numbers of speakers or functional areas of use. At the other end are

languages that are on the verge of extinction. In between are many degrees of greater or lesser endangerment. The *Ethnologue* does not attempt to identify the level of endangerment of each language listed but does specifically identify those languages at the far end of the scale by indicating 'nearly extinct' at the end of the language entry. A language is listed as nearly extinct when the speaker population is fewer than 50 or when the number of speakers is a very small fraction of the ethnic group. In the 15th edition, 497 languages are so designated.

How to identify the level of endangerment of the remaining languages that are not designated as nearly extinct is not necessarily clear. Linguists seek to identify trends in language use, such as a decrease in the number of speakers or a decrease in the use of the language in certain domains or functions. An increase in bilingualism, both in the number of bilinguals and in their proficiency levels, is often associated with these trends. When data are available, the following factors that may contribute to endangerment are reported in the language entries: small population size, bilingualism, urbanization, modernization, migration, industrialization, the function of each language within a society, and whether or not children are learning it. Such factors interact within a society in dynamic ways that are not necessarily predictable. As a scholarly consensus forms that can be applied worldwide, a scale of endangerment is becoming increasingly possible. In the meantime, only brief statements about the above factors are given for each language as data becomes available.

Overview of Contents

The *Ethnologue* begins with an 'Introduction' and 'Statistical Summaries.' The latter give a summary view of the world language situation in terms of numerical tabulations of living languages and number of speakers by world area, by language size, by language family, and by country. Then follows the main body of the work in 'Part 1: Languages of the World.' This section provides detailed information on all known living languages of the world organized by area and by country. An extensive bibliography is located at the end of this section. 'Part 2: Language Maps' provides 208 color maps locating the languages in most countries of the world. Finally, 'Part 3: Indexes' consists of three indexes: a language name index listing all of the 39 491 distinct names that are associated with the languages described in

Part 1, a language code index listing all of the three-letter language identification codes that are used in Part 1, and a country index listing the pages on which the section for that country begins in Part 1 and Part 2.

In Part 1, languages are listed by country under the five major geographic areas of Africa, the Americas, Asia, Europe, and the Pacific. The entry for a country begins with a header giving summary information about the country including official name, total population, a listing of national or official languages, a listing of recent immigrant languages, country literacy rates, and a count of languages indigenous to the country. This header is followed by an entry for each language of the country that is not a recent immigrant.

Entries are alphabetized by the primary name of the language. This is followed by all known alternate names and the three-letter identification code. An estimate of the speaker population is then given; there may also be an estimate of monolingual speakers, or of the size of the ethnic group (including those who no longer speak the language). Next, the location of the language group within the country is described, followed by the genetic classification of the language. Where dialects are known to exist, these are listed along with alternate names. Comments on intelligibility and similarity relationships among dialects or with neighboring languages may follow. Next are notes on language use, including functions of the language (such as official language or language of wider communication), viability, domains of use, age range of speakers, attitudes toward the language, and bilingual proficiency in other languages. These are followed by notes on the status of language development, including literacy rates, use in elementary or secondary schools, scripts used for writing, existence of published literature, and use in media. The entry closes with information in miscellaneous categories including general remarks, linguistic typology, geological and ecological environment, subsistence type of the speakers, and religions.

Bibliography

- Gordon R G Jr (ed.) (2005). *Ethnologue: languages of the world* (15th edn.). Dallas: SIL International. Web edition at <http://www.ethnologue.com>.
 Hale K, Krauss M *et al.* (1992). 'Endangered languages.' *Language* 68(1), 1–42.

Etruscan

D Ridgeway, University of Edinburgh, Edinburgh, UK

© 1994 Elsevier Ltd. All rights reserved.

The Etruscans were an indigenous people of pre-Roman Italy. Although their language has early lexical and other debts to the Indo-European Italic languages, it does not itself belong to the Indo-European group. Written Etruscan is attested from the beginning of the seventh century BC by ca. 13 000 inscriptions, mostly of the fifth–second centuries BC and concentrated in Etruria proper, between the Tiber and the Arno; others confirm Etruscan expansion into Latium, Campania, and Northern Italy. The various types of Etruscan alphabet are derived from that brought to Italy in the eighth century BC by Greek traders, with modifications occasioned by the pre-existing phonetic systems characteristic of different Etruscan-speaking areas. In spite of the popular misconception that etymology and decipherment are still relevant approaches, Etruscan can be read and substantially understood.

Short formulaic inscriptions on tombs or sarcophagi make up the largest category of Etruscan texts; many others define the artefacts on which they appear, often identifying them further as the property or votive offerings of named individuals. Etruscan or Etruscanized Greek names frequently accompany figures in the mythological and other scenes painted in chamber tombs or engraved on bronze mirrors; a few longer texts contain ritual or contractual prescriptions. No Etruscan literature has survived, and only the quasibilingual set of three gold tablets discovered in 1964 at Pyrgi register a historical event (in Etruscan and Phoenician): the dedication ca. 500 BC of a shrine to a Phoenician goddess by the local Etruscan ruler. Exclusive of divine and other proper names, surviving Etruscan vocabulary

amounts to roughly 250 words. The investigation of Greek lexical and onomastic additions to Etruscan has shed light on the development not only of Etruscan phonology but also of the cultural connections of the Etruscans themselves. The statistical treatment of personal names has likewise shown that the addition of gentilicia (incorporating the morpheme *-na* to denote belonging) coincides with the rise of large urban centers. Although inevitably incomplete, systematic descriptions of Etruscan grammar have been prepared on traditional lines (Bonfante 1983). The application of more advanced methods (Cristofani 1979) is inhibited by the brevity and repetitive content of most of the texts, which permit little more than the recognition of adjectives, adverbs, conjunctions, and numerals, and of the inflected nouns, pronouns, and verbs commonly encountered even in short texts on tomb offerings: *mini alicē Velthur* ('Velthur gave me'); *mi Spuriei si Teithurnasi aliku* ('I was given to Spurie Teithurna').

The study of Etruscan must always depend on the comprehension of the Etruscan archaeological and historical context. Ironically, this is wholly lacking for the longest text, a liturgical calendar of the late second century BC now in Zagreb: nothing is known of the circumstances in which the linen book-roll containing its ca. 1200 words became available for recycling into bindings for a mummy, bought in Egypt by a Croatian traveller ca. 1850.

Bibliography

- Bonfante G & Bonfante L (1983). *The Etruscan Language: An Introduction*. Manchester: Manchester University Press.
- Cristofani M (1979). Recent advances in Etruscan epigraphy and language. In Ridgeway D & Ridgeway F R (eds.) *Italy before the Romans*. London: Academic Press.

Europe as a Linguistic Area

T Stolz, Universität Bremen, Bremen, Germany

© 2006 Elsevier Ltd. All rights reserved.

Historical Overview

Areal linguistics of Europe is currently experiencing a period of intensive empirical research and methodological discussion, as the impressive series of

eight volumes resulting from the EUROTYP project (Bossong and Comrie, 1998–2003) and other publications either directly (Bernini and Ramat, 1992) or indirectly (Mayerthaler *et al.*, 1993) related to it amply document. Among their early predecessors in the 19th century, we find the German Indo-Europeanists Schleicher (1850) and Pott (1868; 1887), whose interest in the linguistics of Europe, however, was not primarily guided by purely geolinguistic hypotheses

but rather by the prominence given to questions of language genealogy induced by the *Zeitgeist* of their epoch. Therefore, their studies highlighted diversity rather than convergence – an attitude that continued to dominate well into the 1960s among scholars studying the European linguistic landscape. In the mid-20th century, Lewy ([1942] 1964), most probably inspired by the ‘discovery’ of the Balkan *Sprachbund* by Sandfeld (1930) and the large phonological areas proposed by prominent members of the Prague Linguistic Circle in the 1930s (Jakobson, [1931] 1971), looked at the areal makeup of Europe, mostly in the realm of morphology and morphosyntax. Lewy’s fundamental ideas are also at the heart of the contributions by Becker (1948), Wagner (1959; 1964) and Décsy ([1973] 2000), who, in a manner of speaking, are the representatives of a still somewhat speculative and impressionistic approach with strong affinities to *Völkerpsychologie*, the formerly widespread branch of cultural anthropology that, anticipating Whorf ([1941] 1956) in a way, postulated a high degree of cognitive relativism supposedly determined by and identifiable on the basis of linguistic structure. Becker (1948) turned his attention to the overall convergence of European languages, while Décsy ([1973] 2000) tried to combine the identification of distinct subareas (= diversity) with the deduction of pan-Europeanisms (= convergence), cf. below.

The three booklets by Haarmann (1976a; 1976b; 1977) mark the transition from heavy speculation to a more rigorous methodology mostly rooted in quantitative typology, whereas Bechert ([1981] 1998), evidently impressed by the progress areal linguistics had been making in the description of the linguistic geography of regions outside of Europe, instigated an entire research program for the investigation of Europe as a linguistic area, thus preparing the ground for EUROTYP (König and Haspelmath 1999). More recently, a new approach, meant as a kind of transnational philology, has been developing. It goes by the name of *Eurolinguistik*, and, apart from the clearly cultural–political mission the major representatives have formulated, adopts a pan-European vantage point and attempts to identify those linguistic phenomena that attest to a kind of common European-ness of the languages and speech communities involved. A recent example is Hinrichs’s (2004) collection of articles that address the issue of whether the languages of Europe are currently shifting from the synthetic morphological type to the analytic type. For other linguistic endeavors with a continent-wide orientation, the reader is referred to the survey in Ureland (2004).

There is thus, an abundance of studies dedicated to the description and evaluation of the geolinguistic

situation in Europe. Indeed, these studies all contribute to our knowledge of the areal linguistics of the continent. However, they stem from a variety of schools of thought with at times widely diverging theoretical axioms. This diversity on the theoretical side, of course, has methodological consequences that, in turn, also determine which picture of areality results from the interpretation of the empirical facts.

Approaches

In addition to the commented inventories of European languages without explicit or dominant areal-linguistic goals (Nocentini, 2002; Banfi and Grandi, 2003), there are several approaches to the languages of Europe that all have been adopted at least once in the history of areal linguistics of Europe. The egalitarian approach, the segregating approach, and the center vs. periphery approach are each discussed separately, with these terms being coined here as handy labels for the present occasion. Before we proceed to this survey, some remarks of a methodological nature are in order that affect not just European areal linguistics specifically but areal linguistics in general (including subareas, etc.). For a more detailed discussion of these and similar problems, see the contributions in Matras *et al.* (2005).

With a view to establishing whether or not the distribution of linguistic phenomena over languages has or needs an areal explanation, it is absolutely necessary to clarify in advance a number of difficult issues. First of all, one has to start from somewhere, that is, one has to decide whether the area-to-be has a geographic, political, historical, cultural, or other foundation. Talking about Europe as a linguistic area presupposes an idea of some kind of what is meant by ‘Europe.’ If the starting point is geography, problems arise, as Europe is notoriously difficult when it comes to determining, for instance, where it ends and Asia begins and vice versa. A frequent victim of these uncertainties about the borders separating the two continents is the Caucasian region, which is often totally or partially (= Trans-Caucasus) ignored (Décsy, [1973] 2000), whereas languages such as Georgian, (Eastern) Armenian, Azeri (South Azerbaijani), and so on are classified as ‘European’ in other publications (Siewierska, 1998). A perhaps rhetorical question: If the Caucasus and/or Ural mountain ranges are considered topographic obstacles that supposedly render communication across the line difficult, why should this be different in the case of the Alps? As a matter of fact, individual researchers arbitrarily stipulate certain relatively time-stable landmarks as outer boundaries of the continent, even if one and the same language is spoken on both sides of

the landmark. One example is the Bosphorus, which cuts the Turkish territory into two parts of very different size, the smaller one belonging to Europe, while the larger one forms part of Asia. Does it make sense at all for areal linguistics to follow the lead of geography, especially if the solutions of the latter are inconclusive or variable? These varying solutions do not only have implications for the sample size but also for the general inventory of possible structural properties of European languages and the quantity and quality of potential subareas.

Similarly, taking political boundaries as a yardstick creates more problems than it solves, although it must be acknowledged that they may become relatively strong secondary factors in the reshaping of linguistic boundaries, especially dialect boundaries (Auer, 2004). Since state boundaries do not automatically map onto topographic landmarks of the above-mentioned kind, a political definition of 'Europe' yields a result markedly different from a geographic definition. Under the premise that the notion of 'Europe' is not reduced to the present member states of the European Union, Turkey may again serve as an example of the potential problems a policy-based definition is bound to create: If the entire territory of the Turkish state has to be taken into account, this also means that those languages that are coterritorial with Turkish, such as Kurdish, become European languages. Varieties of Kurdish, however, are also spoken in the adjacent countries of the Middle East – a fact that makes them Asian languages – and thus the political boundary induces the linguist to investigate only part of the territory occupied by a given speech community.

What is more, political boundaries are of course subject to minor and major adjustments according to the momentary balance of power among competing political entities. In other words, the supposed European linguistic area based on political boundaries is prone to metamorphosis in time, as its political basis is in no way immutable. Thus, if we go back a couple of centuries and look at the political boundaries of the *de jure* predecessor of contemporary Turkey, the Ottoman Empire, the option of including or excluding it from the study of Europe as a linguistic area in say, the 17th century, is tantamount to deciding whether to have a sizable part of the Middle East and North Africa in or out of the sample. If the answer is to exclude, then the entire Balkans and other regions and their languages have to be excluded as well because they were under Ottoman rule at that time – a solution that leaves room only for a rather small-sized Europe.

Equally unhandy is the requirement of a common cultural background of those speech communities whose languages belong to one and the same

linguistic area (Décsy, [1973] 2000). One would have to define a culture area beforehand and thus work on the hypothesis that culture areas and linguistic areas are largely coextensive with each other (Lewy, [1942] 1964; Becker, 1948). Haarmann (1976b: 71–76) demonstrates that the culture-first approach rests on the erroneous assumption that cultural convergence and linguistic convergence are two sides of the same coin, which is evidently much too strong a hypothesis. Apart from the fact that it would surely cause problems to depict Europe sweepingly as culturally homogeneous, it must be stated that cultural traits and linguistic traits may diffuse over geographic boundaries, but not necessarily in a parallel fashion. On the other hand, neither of the two kinds of traits automatically covers a geographically defined region in its entirety. Thus, instead of starting from mostly elusive or controversial nonlinguistic criteria, the more promising approach to linguistic areality is the one suggested by Bechert ([1981] 1998), according to which the diffusion patterns of the individual linguistic phenomena, independent of geographic, cultural, and political assumptions, should guide the linguist. This is what generations of dialectologists have successfully been doing when they identify isoglosses and define dialects as those varieties that share a certain set of isoglosses (or isopleths, i.e., clusters of isoglosses) (van der Auwera, 1998).

Equally important for the outcome of one's research is the status of the sample languages. The exclusive comparison of written varieties of normative standard languages will inevitably yield results different than an approach that takes nonstandard varieties (regional languages, dialects, etc.) into account (Haarmann, 1986). Recent research emphasizes the fact that we need to pay more attention to substandard varieties in order to better understand – among other things – the areality of linguistic phenomena (Kortmann, 2004). With these provisos in mind, we are now in a position to tackle the issue of the coexistence of various approaches.

The Egalitarian Approach

The egalitarian approach presupposes that all languages that qualify as European share certain features and thus display a sufficient degree of similarity, such that the entire geographic space occupied by these languages can be considered a linguistic area. Therefore, this approach adopts *per se* a continent-wide perspective. The most radical version of the egalitarian approach can be found in Becker (1948), who treats the languages of Europe as a solid homogeneous block. The evidence for this is, however, rather poor and consists mainly of idiomatic or

phraseological parallels with a more or less obvious origin in Christian religious discourse practices. Décsy ([1973] 2000) is slightly more cautious, as he also recognizes that Europe is fragmented into several subareas and thus that there is a potentially rather high degree of heterogeneity of the continent. This acknowledged internal diversity notwithstanding, Décsy ([1973] 2000) searches for pan-European commonalities that he terms ‘Europemes.’ Among these Europemes, we encounter, for instance, a kind of common core phonological system with five vowels (= /a/, /e/, /o/, /i/, /u/) and 10 consonants (/p/, /t/, /k/, /s/, /v/, /m/, /n/, /l/, /r/, /j/) that are said to be present in each and every phonological system in Europe – a somewhat dubious assumption. For reasons of space, the tenability of Décsy’s hypotheses is not reviewed here (but see Stolz, 2002). Since the identification of full-blown Europemes (in the sense of exceptionless substantial areal traits) on all linguistic levels is, for various reasons, next to impossible, Décsy ([1973] 2000: 341) resorts to a solution reminiscent of Lewy ([1942] 1964: 103–108) and Becker (1948) when he says:

[t]he genuine resemblance [of the languages of Europe] lies however in the way of thinking [...]. [T]he great majority of Europeans hold the same mind-set, which arose from a Graeco–Latin cultural tradition in a two-thousand year development, connecting the regions between Iceland and the Urals, between the North Cape and Palermo, between Dublin and Istanbul to a unity in historical proportion.

This is equivalent to a declaration of defeat in linguistic terms because it admits that one cannot find sufficient tangible proof of a pan-European linguistic structure.

Haarmann (1976a: 105–106) adopts the notion of Europeme but defines it more rigorously in analogy to the well-established Greenbergian probabilistic universals. Haarmann (1976a: 107–108) goes on to sketch the methodological and empirical requirements necessary for a wholesale comparison of the languages of Europe. On a still provisional basis of selected aspects of 65 sample languages excluding the Caucasian region (the ultimate goal being the large-scale comparison of entire grammatical systems), Haarmann (1976a: 108–116) puts forward 16 Europemes for which he feels to have sufficient evidence (the following Europemes implicitly contain the common introductory phrase ‘In all European languages,’ or a suitable variation thereof; where the Europemes come in the shape of an implication, it is always meant to be unilateral):

- Europeme 1: The number of simple phonemes ranges between 10 and 110.

- Europeme 2: Consonants outnumber vowels.
- Europeme 3: Only one-third of the potential phonotactic combinations is actually made use of.
- Europeme 4: The basic syllable structure is (C)V(C) (C).
- Europeme 5: The basic morphotactic structure is [radical](-[derivation]) (-[inflection])
- Europeme 6: Singular and plural are distinguished formally.
- Europeme 7: Nouns and verbs are distinguished formally.
- Europeme 8: Present and past tense are distinguished formally.
- Europeme 9: Morphological case distinctions range from zero to 30.
- Europeme 10: Indicative, imperative, and conditional are distinguished formally.
- Europeme 11: Multifunctional derivational affixes are ubiquitous and outnumber monofunctional ones.
- Europeme 12: Synthetic and analytic encoding strategies are compatible with each other.
- Europeme 13: Changes from synthetic to analytic structures do not necessarily embrace the whole grammatical system.
- Europeme 14: The order $S > O$ is basic (V may occupy various positions).
- Europeme 15: VSO languages employ the category of verbal nouns.
- Europeme 16: Intonation contours in questions are different from the ones used in declarative sentences.

Some of the above generalizations are almost trivial because they meet universal expectations. Others are not particularly distinctive, and still others remain somewhat vague as they circumscribe the range of variation between a minimum and a maximum number of entities. Already Europeme 1 is problematic, as the upper limit is imposed by Kildin Saami, with whose large phoneme inventory of over 100 none of the other European languages can compete (there is a gap of over 40 units between Kildin–Saami and the second best, Lithuanian). At the same time, it remains opaque which of the languages of Haarmann’s sample comes close to the minimal size of phoneme inventories, as Haarmann (1976a: 113) himself states that European languages generally do not display significantly small inventories. Similarly, the rich morphological case-systems suggested by Europeme 9 boil down to Hungarian, which, according to the maximalist count, displays 28 cases and thus exceeds the paradigms of its competitors (Finnish, Basque) by about a dozen units. Instead of working with the

statistical extremes, it would probably have made more sense to go by the statistical mean or any other mathematical procedure on which predictions about the expected average system can be formulated. The Europemes allowing for quantitative variation therefore can be considered to be merely observations of structural facts that cannot claim to identify a certain quality that characterizes the languages of Europe. In addition, Europeme 13 is odd in the sense that it is the only diachronic generalization in a list of otherwise strictly synchronic observations, and Europeme 15 has a genetic bias as it only applies to members of the Celtic phylum.

No matter how accurately the Europemes capture the linguistic facts, Haarmann's (1976a) approach works in such a way that the languages of his sample are automatically treated as similar in one way or another (which is a pitfall of the geography-first approach criticized by Haarmann (1976b) himself). This is for methodological reasons: The researcher focuses on the common denominators alone without specifying whether or not the findings have to be explained by diffusion. Any two languages of the world, past and present, can be compared in order to determine what they have in common. Whether the shared properties call for a genetic, typological, areal, or other explanation, if at all, is a completely different matter. Thus, the above Europemes gain significance for areal linguistics if and only if it can be demonstrated that they result from diffusion (either way, i.e., receding or expanding) and/or that their combination is characteristic of the area under scrutiny such that it can be kept apart from other areas.

The Segregating Approach

The segregating approach starts from a different premise, according to which not all European languages resemble each other to the same extent; the degree of similarity may be significantly stronger in certain smaller subareas of the continent that, in turn, deserve the designation of distinct linguistic areas of their own. Thus, this approach favors research on individual regions, not necessarily with an orientation toward comparing the findings with those of other potential subareas. Haarmann (1976b), and to a lesser extent Décsy ([1973] 2000), are exceptions as they, in principle, combine the segregating and the egalitarian approaches.

On the basis of a sample of 18 languages, Lewy ([1942] 1964) distinguishes five subareas in Europe. The languages discussed are to be understood as representatives of larger groups forming geographic neighborhoods:

- Atlantic area (Celtic, Romance [without Rumanian (**Romanian**)], East and West Scandinavian, English)
- Central area (German, Hungarian)
- Balkan area (Albanian, Rumanian, Greek)
- Eastern area (Baltic, East Slavic, Finnish, Mari, Mordvin (Erzya))
- Arctic area (Yurak)

With the notable exception of the Balkan area, Lewy's proposals have experienced major alterations in the subsequent years. The changes are partly occasioned by the fact that the number of sample languages increased considerably from 65 in Décsy's and Haarmann's publications in the 1970s to 140 or more for EUROtyp and related projects. The more important factor, however, is the introduction on the list of items to be compared of further categories and phenomena on all linguistic levels. Presently, 12 potential subareas (often called *Sprachbünde* or the like) are discussed with varying intensity in the pertinent literature, even though most of them are still largely controversial and some – like the Littoral *Sprachbund* – are unlikely to pass the test at all. In the following list, the languages that count as members of a given subarea are the ones explicitly mentioned by the first source quoted; in the literature, different sets of languages may occur under the same heading, whereas different labels may be used for the same set of languages. This variation reflects the fact that the shape of a linguistic area is directly dependent on the quality and quantity of the features one scrutinizes (Haspelmath, 2001: 1505).

- Standard Average European (SAE) languages (= Albanian, Dutch, French, German, Italian, Portuguese, Sardinian, Spanish) (Haspelmath, 2001) including the *Charlemagne-Sprachbund* (van der Auwera, 1998: 824); cf. below.
- Viking *Bund* (Décsy, [1973] 2000)/northwest linguistic area (Wagner, 1964) (= insular and mainland North Germanic, Celtic, Saami, Finnish, Veps, and the Anglo-Saxon component in modern English)
- British (Isles) areal type (= Celtic, English) (Wagner, 1959), the relic of an erstwhile more extended area to which also Basque and Berber belonged
- Littoral *Bund* (= Basque, Dutch, Frisian, Maltese, Portuguese, Spanish) (Décsy, [1973] 2000), based on the idea that the speech communities involved represent renowned seafaring nations
- Mediterranean linguistic area (= Romance, Maltese, North African Arabic, Hebrew, Turkish, Balkan *Sprachbund*, Serbian, Croatian, Slovenian) (Ramat and Stolz, 2002)

- Rokytno *Bund* (= Bielarussian (Belarusan), Kashubian, Lithuanian, Polish, Ukrainian) (Décsy, [1973] 2000)
- (Circum-)Baltic super-position zone (Koptjevskaja-Tamm and Wälchli, 2001)/ Peipus *Bund* (Décsy, [1973] 2000) (= Balto-Finnic, Danish, Estonian, Latvian, Lithuanian, Low German (Low Saxon), Saami, Swedish)
- Karelian *Sprachbund* (Sarhimaa, 1991) (= Russian, Balto-Finnic)
- Eurasian *Sprachbund* (East Slavic, conservative Polish and Sorbian, Rumanian, Lithuanian, Altaic languages of the former Soviet Union) (Stadnik, 2002)
- Danube *Sprachbund* (= Czech, German, Hungarian, Slovak) (Skalička, 1968)
- Volga-Kama *Sprachbund* (= Bashkir, Chuvash, Kalmyk (Kalmyk-Oirat), Mari, Mordvin, Tatar, Votyak (Udmurt), Yurak (Nenets), Komi-Zyrian) (Wintschalek, 1993)
- Caucasian *Sprachbund* (= all Caucasian phyla) (Haarmann, 1977)

Décsy ([1973] 2000) does not tolerate a single European language to remain outside a *Bund*, and thus he postulates a number of so-called groups of languages (viz., Isolates [in the sense of 'not belonging to any other *Bund*'] and Diaspora languages) that have no areal *raison d'être* but are said to be based on social and/or historical criteria. Haarmann (1976b) rightly observes that this *horror vacui* of Décsy's renders the whole areal undertaking dubious since the classificatory criteria are not kept constant, which is characteristic of Décsy's handling of the problems posed by European areal linguistics (Russian, for instance, counts as an SAE language just because it happens to have more than 50 million speakers; thus, demographic criteria may oust structural ones for Décsy). Moreover, languages do not have to converge even though they happen to be neighbors. The fact that the territories of two speech communities adjoin surely facilitates convergence but does not necessarily trigger it. Irrespective of these additional methodological problems, the many and sometimes competing suggestions of subareas within and beyond Europe are indicative of a certain linguistic heterogeneity of the continent. This heterogeneity on the meso level and micro level raises the question of whether it makes sense at all to talk about Europe as a linguistic area on the macro level.

Center vs. Periphery Approach

The center vs. periphery approach is not so much a compromise between the two previous ones (which

operate on the basis of predetermined areas with more or less fixed sets of member languages) as it takes a dialectology-minded stance. It assumes that individual linguistic phenomena diffuse geographically in such a way that it becomes difficult to determine clear-cut boundaries between those languages that form part of a given linguistic area and those that remain outsiders. Membership in a linguistic area is thus a matter of degree, a gradient property of the languages compared (Haspelmath, 2001). The approach is feature based with a continent-wide perspective (or more precisely, with a potentially unlimited perspective, as no boundaries are defined beforehand). This approach plays an important role in the recent discussion about the notion of SAE languages.

Haspelmath (2001: 1493–1501) presents a list of 12 properties that he classifies as major SAE features, all of which belong to the realm of morphosyntax in the widest sense of the term. In the subsequent list of features, the abbreviation 'SAE' stands for varying sets of languages that nevertheless overlap in such a way that some languages partake in (almost) every SAE-isogloss (namely French and German, the two pillars of the *Charlemagne Sprachbund*):

- presence vs. absence of definite and indefinite articles: SAE languages have both types of articles, whereas languages on the fringes of and beyond the SAE area either lack one or both.
- relative pronoun strategy in relativization: SAE languages employ relative pronouns, i.e., relativizers that contain grammatical information about the syntactic function of the relativized head in the relative clause, whereas outside the SAE area, languages opt for invariable relative particles or other strategies.
- 'have'-perfect: In SAE languages, an auxiliary with the (erstwhile) lexical meaning of 'have' serves the purpose of encoding perfect (sometimes in complementary distribution with 'be'-perfects); non-SAE languages prefer 'be'-perfects or other strategies.
- agent-like experiencers: SAE languages tend to encode experiencers as subjects and thus treat them on a par with agents, whereas languages in the European east and also in the extreme northwestern regions have a predilection for keeping agents and experiencers formally apart; this difference in semanto-syntactic behavior is, however, a matter of degree, as both agent-experiencer 'syncretism' and agent-experiencer 'differentiation' can be encountered on both sides of the dividing line.
- participial passive: SAE languages have a passive construction made up of an auxiliary and a passive participle of the lexical verb, a construction type that is absent from the languages spoken

in the eastern part of the continent and in the westernmost languages.

- anticausative–prominence: Inchoative–causative alternations are expressed predominantly by anticausatives in SAE languages, whereas causatives are preferred by languages outside this area; as with agent-experiencer encoding, we are dealing with a preference for anticausatives and not with a 100% solution.
- dative external possessors: SAE languages typically have external possessors and encode them as datives, which contrasts with external possessors of the locative type found in North Germanic and with languages that lack external possessors altogether.
- negative pronouns/lack of verbal negation: SAE languages normally combine a verb unmarked for polarity with a negative indefinite pronoun, while outside this area, i.e., on the western and eastern rims of the continent, negative pronouns go along with negated verb forms.
- particle comparatives: SAE languages overwhelmingly employ particle comparatives, i.e., the standard of a comparative construction is introduced by a conjunction-like element; in other European languages, the locative strategy of marking the standard is more frequent.
- equative constructions based on the relative clause: Equally in the realm of comparison, equative constructions reveal that SAE languages resort to a kind of adverbial relative clause that contains the standard of comparison; this strategy is unknown in the languages spoken in the west, north, and east of Europe, where special equative markers or other means are employed.
- strict agreement markers: SAE languages inflect their verbs for subject person even though an overt subject NP is always copresent (= ‘non-prodrop’), whereas the vast majority of European languages are of the referential-agreement type (= ‘prodrop’).
- intensifier-reflexive differentiation: SAE languages distinguish intensifiers (= words that mark the referent of an NP as central) from reflexives, whereas nondifferentiation of the two categories is widespread among non-SAE languages.

These findings stem from a variety of studies, the empirical basis of which differs widely because of different sample size and so forth. Some of the features have a wider distribution that embraces the bulk of the languages of Europe, whereas others are attested only in a relatively small number of languages. Some are characteristic of European languages

in general and let them stand out as ‘exotic’ (Dahl, 1990), whereas others are preponderantly Indo-European or characterize only a subset of the languages of Europe. Independent of these differences in their distribution in space, the phenomena have one crucial element in common, viz., they all involve larger or smaller geographically contiguous areas. In other words, they do not occur sporadically on the map but cover entire regions (with the occasional isolate or secondary accumulation outside the more extended area). More often than not, the isoglosses cut across major (= macrophyla such as Indo-European and Uralic, etc.) or minor (= phyla and subphyla such as Germanic and Romance, North Germanic and West Germanic, etc.) genetic groups. This is of course indicative of a possible origin in diffusion via language contact.

The various isoglosses overlap, which allows us to identify a core area with a particularly high number of shared features, as opposed to a periphery in which languages only participate in a smaller number of isoglosses. Ideally, the following generalization holds: The further away one gets from the core, the smaller the number of shared features. In a manner of speaking, a given language may be more or less SAE according to the number of isoglosses in which it partakes. Being an SAE language is thus again a matter of degree. However, this observation cannot solve all the problems acknowledged by Haspelmath (2001: 1504–1506). Apart from the fact that the complete absence of shared features may be taken as evidence of a language’s areal outsider status, it is the linguists’ choice of features that has a bearing on the identification of areas. A different catalogue of properties might yield a completely different geolinguistic map of the same group of languages. Moreover, there remains a certain element of arbitrariness when it comes to deciding what the numbers of attested isoglosses tell the observer. As a matter of fact, it is up to the individual linguist to decide how many features are required for a language to be a member of the core or the periphery – and, on top of that, whether the number of shared features is in any way significant. For the solution of the latter problem, comparative studies of linguistic areas worldwide are called for.

Nevertheless, the validity of the SAE area is well documented and based on empirically solid foundations. In addition to the above major features, Haspelmath (2001: 1501–1504) mentions many further candidates for the status of SAE isogloss; however, their exact geolinguistic distribution has not been established yet. Some of these additional features provide evidence *ex negativo*, as they state the absence of a given feature in SAE languages (lack

of grammatically relevant alienability correlations in adnominal possession; lack of formal inclusive–exclusive distinction on pronouns; lack of (partially) reduplicating constructions). Three of these supposedly minor features are discussed below, based on research carried out by the present author. While one may agree with Haspelmath that negative evidence has to be taken into account when it comes to identifying linguistic areas, it is necessary to also accept the methodological consequences. If the absence of certain features is characteristic of SAE languages and is thus counted as an isogloss defining a linguistic area, then the absence of typical SAE features in languages outside the SAE area may likewise be considered evidence *ex negativo* in favor of a linguistic area (especially if the features in question are in a binary nongradient relation). Thus, a strong competitor of the SAE *Sprachbund* is created as an epiphenomenon of our search for isoglosses supporting the SAE hypothesis. It would certainly be unfortunate to baptize this additional area the ‘non-SAE area’ (not simply because it might ultimately turn out to comprise the majority of the languages of Europe). If one wants to deny those languages that lie outside the SAE *Sprachbund* the status of an area of their own, it is a necessity to find strong counterarguments, among which only internal heterogeneity seems to be convincing.

The center vs. periphery approach reveals certain important facts. In addition to the idea that membership in a linguistic area is a gradient property, Haspelmath’s findings also suggest that one and the same language may belong to several different linguistic areas at the same time (for example, a marginal member of the SAE *Sprachbund* might be a marginal or central member of the still nameless non-SAE competitor). Given that the linguist’s choice of features has a considerable influence on what a linguistic area might look like, there is at least theoretically the possibility that the prominence given to SAE languages in the extant literature does not adequately reflect the areal composition of the continent. Chances are that the members of the SAE *Sprachbund* identified so far behave differently as soon as we take other linguistic phenomena into consideration, thus relativizing the current emphasis on the SAE *Sprachbund* by showing that there are potentially other areas in Europe that could pass as serious competitors. Haspelmath (2001: 1505–1506) acknowledges these possibilities, though with certain reservations. In the next section, with a view to elaborating on these issues, a few phenomena are scrutinized that either have been ignored completely or mentioned only in passing in the recent discussion within European areal linguistics.

More Isoglosses

The evidence adduced by Haspelmath is exclusively morphosyntactic; below, there is a variety of phonological issues that also display a clear geographical distribution. For the sake of brevity, another good candidate for areality in the realm of phonology, i.e., Sandhi, cannot be discussed here (Andersen, 1985). In addition, areal phenomena in Europe are approached from the perspective of grammaticalization theory by Heine and Kuteva (2005). The phonological sketches are complemented by others that focus on morphology. The list of features is again random, apart from the fact that they are not discussed in detail in Haspelmath (2001). When reference is made to more or less central languages or ‘layers’ of SAE, these judgements are based on Haspelmath’s (2001: 1505) cluster map meant to represent the degrees of membership in the supposed linguistic area. The following (convenience) sample comprises 51 languages of all macrophyla represented in Europe (= Indo-European, Uralic, Altaic, Caucasian, Afroasiatic [= Maltese] and the isolate Basque). For the sake of simplicity, exclusively data from standard varieties are compared (occasionally allowing for a comparison of competing standards). In the east, the dividing line between Europe and Asia is assumed to run from the Arctic Sea southward along the Ural mountains to the Caspian Sea, then westward along the southern borders of Azerbaijan, Armenia, and Georgia to the Black Sea, where it follows the Turkish state boundary to the Mediterranean; in other words, the entire Trans-Caucasus and Anatolia are treated as integral parts of Europe. For the northern, western, and southern limits of Europe, a conventional solution has been taken.

Phonology

Rounded front vowels The first feature to be scrutinized is the presence or absence of vowel phonemes with the feature triple [–low] + [front] + [round] = /y/, /Y/, /ø/, /œ/ as in French *culture* [kylytʁ] ‘culture’ and *sœur* [sœr] ‘sister.’ The areal distribution of rounded front vowels is also partly discussed in Chambers and Trudgill (1980). Eighteen languages of the sample used here have two rounded front vowels, and one rounded front vowel is attested in two further languages, leaving a solid majority of 31 languages that lack vowel phonemes of this kind. **Figure 1** reveals that there are two hotbeds in Europe where rounded front vowels occur.

These hotbeds are at a considerable distance from each other and are separated by a solid block of languages that do not have these vowels on their

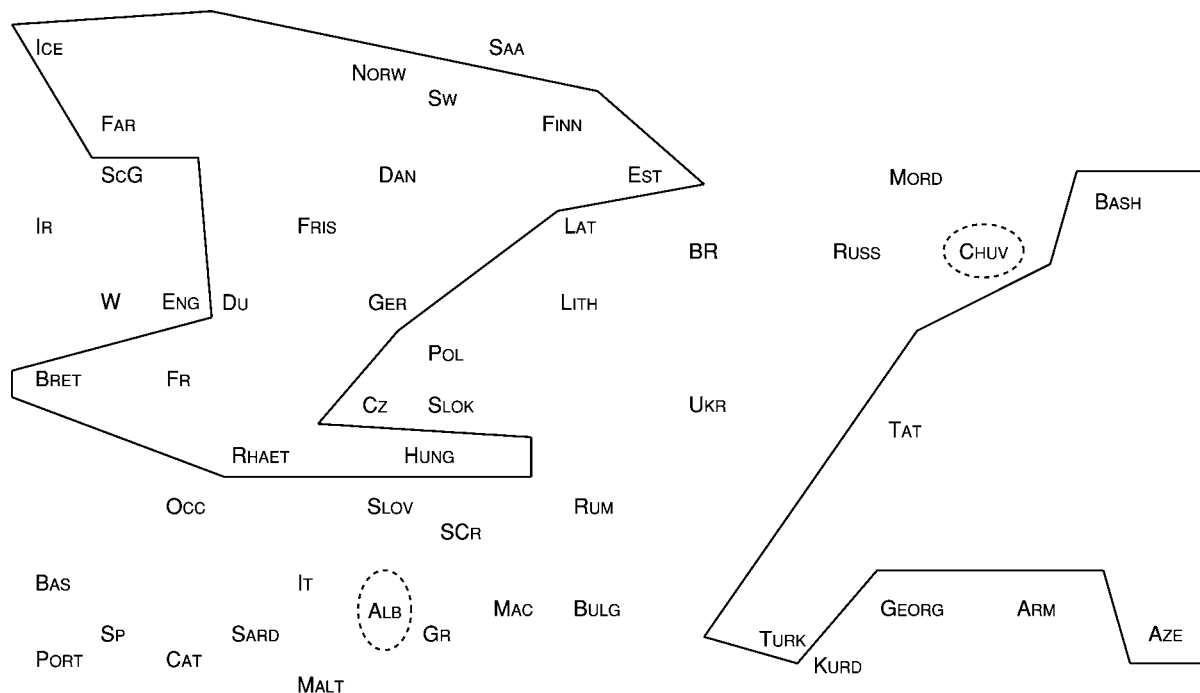


Figure 1 Rounded front vowels [solid line = both /œ/ ~ /ø/ and /ɤ/ ~ /ɥ/; dotted line = only /ɤ/ ~ /ɥ/]. ALB = Albanian, ARM = Armenian, ART = article, AZE = Azeri, BAS = Basque, BASH = Bashkir, BR = Bielarussian, BRET = Breton, BULG = Bulgarian, CAT = Catalan, CHUV = Chuvash, CZ = Czech, DAN = Danish, DU = Dutch, ENG = English, EST = Estonian, FAR = Faroese, FINN = Finnish, FR = French, FRIS = Frisian, GEN = genitive, GEORG = Georgian, GER = German, GR = Greek, HUNG = Hungarian, ICE = Icelandic, IR = Irish, IT = Italian, KURD = Kurdish, LAT = Latvian, LITH = Lithuanian, MAC = Macedonian, MALT = Maltese, MORD = Mordvin, NORW = Norwegian, OCC = Occitan, PERF = perfective, PL = plural, POL = Polish, PORT = Portuguese, RHAET = Rhaeto-Romance, RUM = Rumanian, RUSS = Russian, SAA = Saami, SAE = Standard Average European, SARD = Sardinian, ScG = Scots-Gaelic, SCR = Serbo-Croatian, SLOK = Slovak, SLOV = Slovene, SP = Spanish, SW = Swedish, TAT = Tatar, TURK = Turkish, UKR = Ukrainian, W = Welsh.

phoneme charts. Those languages that have rounded front vowels in the east belong to the Altaic macro-phylum (more precisely, to the Turkic phylum), and thus this feature most probably has a genetic explanation. In the other hotbed, however, Indo-European languages of three different phyla (Germanic, Romance, and Celtic) and Uralic languages share the feature. However, not all members of the various phyla partake in this isogloss. French and the Puter (Lower Engadine) variety of Rhaeto-Romance are the only Romance languages to display rounded front vowels; the same holds for Breton among the Celtic languages. English, however, does not follow its Germanic relatives, all of which have rounded front vowels. Likewise, Saami and Mordvin, both lacking these phonemes, do not go along with their Uralic relatives. The group of languages that make do without rounded front vowels is genetically heterogeneous, too. Note, however, that the entire Slavic and Baltic branches are immune against rounded front vowels in a manner of speaking. Nevertheless, the isogloss cuts across genetic boundaries. The languages that share the feature are geographical neighbors: Those that have rounded front vowels are

spoken next to each other, and those that do not employ rounded front vowels are also in a neighborhood relation among each other. We do not encounter islands of either type interspersed among languages of the opposite type. However, there are two special cases, Albanian and Chuvash, which have /ɤ/ as a phoneme but lack mid-high rounded front vowels and thus fail to partake in the larger isoglosses. Chuvash is renowned for its at times idiosyncratic behavior, as opposed to other members of the Turkic phylum. In the present case, Chuvash seems to mark the transition from the Turkic solution with two rounded front vowel phonemes to the situation found in the surrounding non-Turkic languages (Mordvin, Russian, etc.) where this class of phonemes is lacking, i.e., the absence of /ø/ is likely to be a product of language contact between Chuvash and its genetically unrelated neighbors. The complete absence of rounded front vowels in Mordvin has perhaps a similar explanation. Still, Chuvash is located next to the hotbed of rounded front vowels in the east containing the sister languages of Chuvash. Albanian, on the other hand, is cut off from the hotbed in the west by members of the Slavic phylum that do not

allow for rounded front vowels at all. In all likelihood, Albanian is the remnant of an erstwhile more extended subarea to which, on earlier stages, Greek also belonged. Similarly, English and Welsh have lost rounded front vowels in the course of their history. For Breton, French, and Rhaeto–Romance, the feature is certainly a relatively late contact-induced acquisition: The two Romance languages have copied the feature from Germanic, and Breton developed it under French pressure. It is noteworthy that the *Charlemagne Sprachbund* is again involved, while other more central SAE languages are somewhat underrepresented.

Quantity This section is concerned with the distribution of phonemic length in vowels (= /V:/), as in German *Hasen* [ha:zən] ‘hares’ vs. *hassen* [ha(s)sən] ‘to hate,’ and consonants (= geminates /K_iK_i/), as in Italian *sanno* [sanno] ‘they know’ vs. *sano* [sa(ː)no] ‘healthy.’ Ternes (1998) has studied both phenomena in a survey of the phonology of European languages that tacitly corrects a number of hypotheses put forward by Haarmann (1976a). Owing to the fact that, in a variety of languages, quantitative differences epiphenomenally go along with more or less significant qualitative differences and are sometimes determined by the moraic prerequisites of the canonic syllable structure of a given language, it becomes

difficult to decide whether length itself is the phonemically relevant factor. This and other aspects pose serious problems *inter alia* for the analysis of English and French. In the latter language, vowel length is marginally phonemic only in careful educated speech. In order to keep the discussion within reasonable limits, things are simplified by leaving problems of vowel quality and moraic templates aside. Moreover, cases like the French one are treated as instances of absence of a vowel quantity correlation.

Figure 2 surveys the presence and absence of phonemic long quantity for both vowels and consonants in Europe. Twenty-five sample languages (= slightly less than 50%) have distinctive vowel length, whereas only six display a quantity correlation for consonants. Four of these latter five distinguish quantities for both vowels and consonants. Were it not for Italian and Sardinian, one could be tempted to formulate an implication according to which phonemic quantity of consonants implies phonemic quantity of vowels. Long consonant phonemes are clearly a minority solution.

Independent of the fact that vowel length is distinctive in almost half of the present sample languages, the languages where this feature is attested display a clear areal bias. Discounting the isolated instance of distinctive vowel length in Maltese, one immediately notices that phonemic long vowels are a matter of

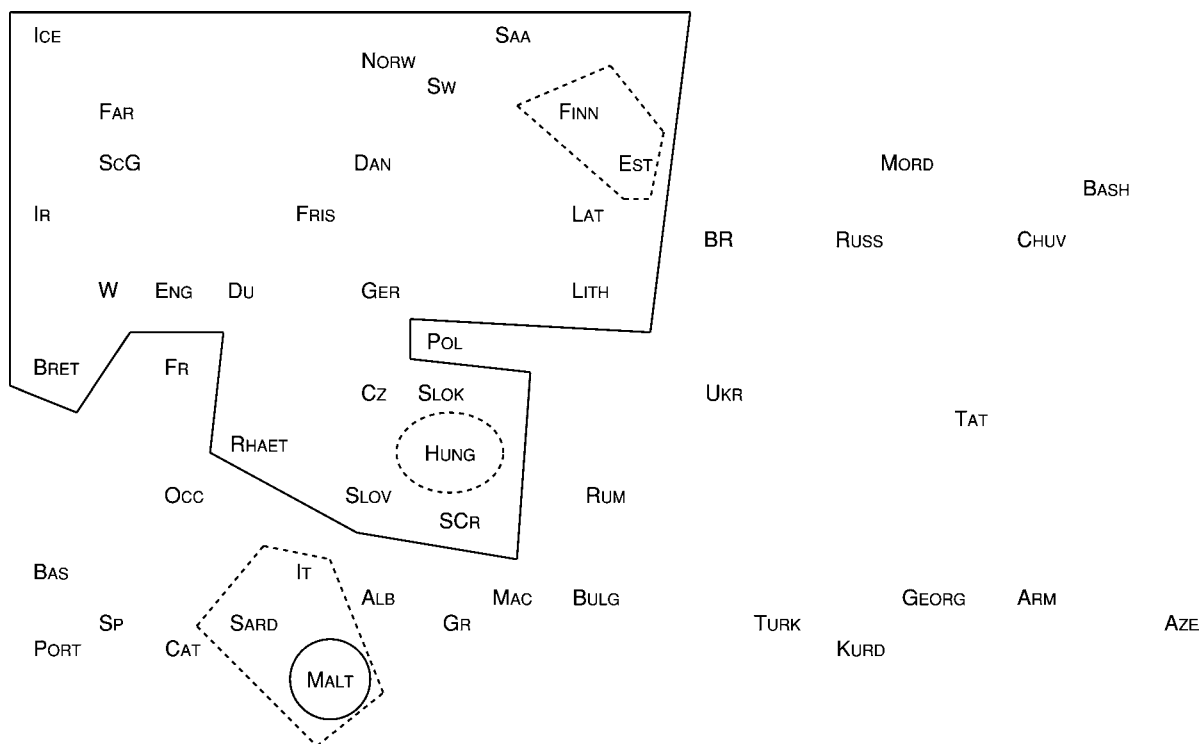


Figure 2 Quantity correlation [solid line = phonemic V:; dotted line = phonemic K:].

a large area comprising the European northwest, Scandinavia, the Baltic region, and central and part of southwest Europe. Three Indo-European phyla in their entirety partake in the isogloss: all Celtic, Germanic, and Baltic languages distinguish short vowels from long vowels. This distinction is also found in varieties of Rhaeto-Romance and some members of the South Slavic and West Slavic subphyla, all of which border the territory occupied by Germanic languages. Of the Uralic languages, those that are spoken in the vicinity of Germanic, Baltic, or Slavic languages with length distinctions for vowels have the same feature, whereas Mordvin seems to lack it. Again, the feature cuts across major and minor genetic boundaries. The same is true of the group of languages that do not allow for phonemic quantity.

The areality of the phenomena under scrutiny cannot be denied. In contradistinction to the foregoing one, where it is possible to identify waves of spread and recession, much seems to speak in favor of a slow but continuous shrinking of the area characterized by phonemic length. Given the fact that, historically, all Indo-European languages employed distinctive vowel quantity, the present situation is suggestive of a large-scale loss of the feature in the successor languages of Latin, Ancient Greek, Old Church Slavonic, and so on. Thus, the area from which phonemic length is absent has been expanding to the detriment of the hotbed of distinctive quantities. Interestingly, the two members of the *Charlemagne Sprachbund* are

this time located on different sides of the dividing line: German is an uncontroversial case of a language with phonemic vowel length, and French does not form part of the isogloss (but see above). With the exception of Dutch, vowel length distinctions are not typical of other more central SAE languages.

Morphology

Morphological case distinctions European languages come in two varieties. One group has case inflections on nouns independent of the exact size of the paradigm (Finnish *talo* {house} 'house' vs. *talo-n* {house}-{GEN} 'of a/the house'), whereas the other group does not mark case on nouns by morphological means (cf. the Welsh 'construct state' *tŷ y dyn* {house} {ART} {man} 'the man's house'). For the present purpose, neither the genitive clitic of English and the mainland Scandinavian languages nor the facultative *s*-genitive on proper nouns in Dutch and Frisian are counted as instances of bound case morphology. The present sample languages distribute over the two possible types as follows: a minority of 20 languages lack nominal case inflections, as opposed to a majority of 31 languages that employ more or less sizable inventories of morphological cases on nouns. **Figure 3** captures the areal pattern.

Nominal case morphology is largely a trait of languages located in the east and in the far west of the continent. A small strip extending from

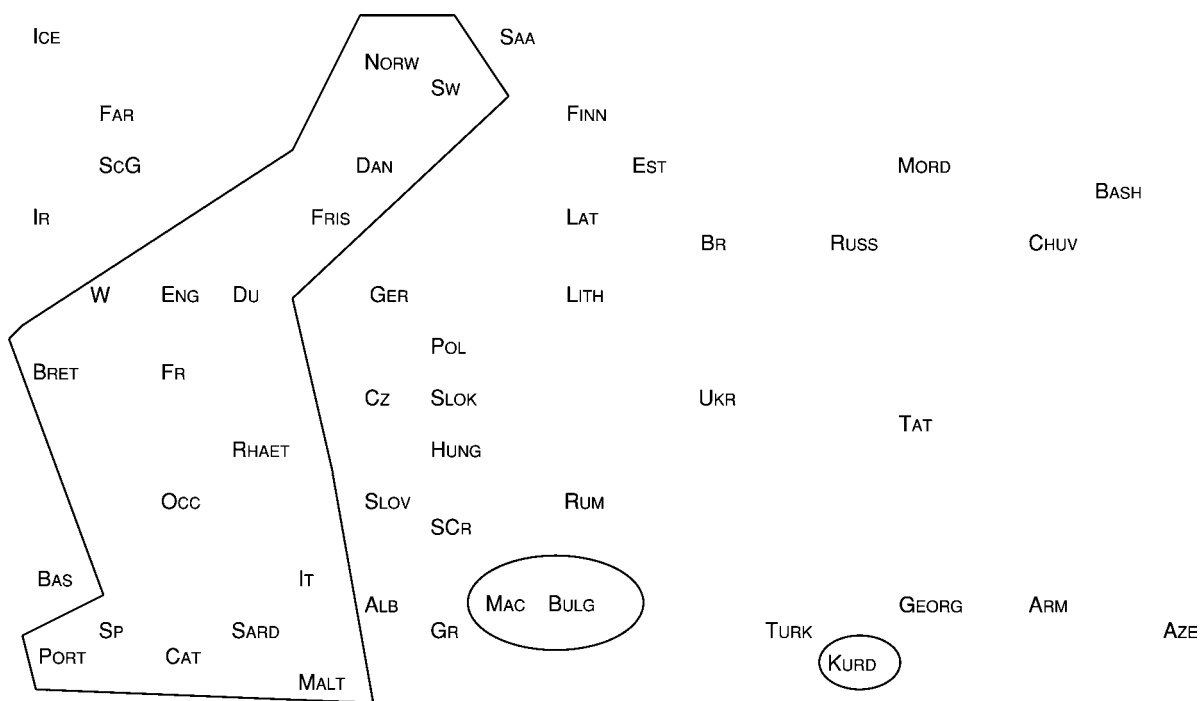


Figure 3 Morphological case on nouns [solid line = languages without case inflection].

Norwegian in the north via French in the middle down to Maltese in the south is characterized by absence of morphological case distinctions on nouns. Outside this area, this feature is also shared by Kurdish in the southeasternmost corner of the map and the two Balkan Slavic languages Macedonian and Bulgarian, which are surrounded by full-blown case languages. Basque is the only case-inflecting island in largely caseless surroundings. As is the situation with the previous isoglosses, phyla are cut across. There are case-inflecting Celtic languages and caseless ones; the same applies to the Germanic phylum, where German and the insular North Germanic languages stand out as representatives of the case-inflecting type, whereas the remaining Germanic languages have lost the ability to inflect nouns for case. Rumanian is the only Romance language with nominal case-inflection, while Bulgarian and Macedonian are dropouts from the Slavic phylum, as they have given up their erstwhile fully functional case system. All Uralic and Altaic languages are case-inflecting, as are the Baltic languages, Armenian, Greek, Albanian, and the bulk of the Slavic phylum. Maltese is the only non-Indo-European language without nominal case inflection. The primary stronghold of case inflection is clearly the eastern part of the continent. In the northwest the feature seems to be on the decline in the Celtic languages, less so in Faroese and probably not yet in Icelandic.

In their older stages, the Indo-European languages (Old Irish, Old English, Latin, Old Church Slavonic, Old Persian, etc.) were all endowed with a nominal case paradigm that in its most conservative shape resembled closely the one of present-day Lithuanian. Therefore, the absence of nominal cases in many Romance, Celtic, Germanic, and some Slavic languages (and Kurdish, for that matter) is an innovation. Much the same can be said of Maltese in comparison to Classical Arabic. In the languages that currently inflect their nouns for case, this feature can be classified as a retention, a conservatism. Old French is reported to have employed a minimal bipartite case system during the earliest stages of documentation, while the nominal case system had already disintegrated in most of its sister languages outside the Balkans. The chronology of the disintegration of nominal case systems is suggestive of a spread of caselessness from the south northward, eventually cutting the formerly continuous area of case languages into two subareas. Presently, the isogloss separates the two members of the *Charlemagne Sprachbund* that again happen to be situated on different sides of the line. The absence of case inflection, however, unites many of the more central SAE languages (Dutch, Spanish, Portuguese, Sardinian, Italian).

Comitative-instrumental syncretism The next feature also belongs to the realm of case distinctions, although it is not restricted to bound morphology. The focus is on the grammaticalized expressions of comitative and instrumental relations independent of the morpheme status of their markers. The presentation is based on the findings of a long-term crosslinguistic research on the two categories, some of which are discussed in Stolz *et al.* (2003), Haspelmath (2001), and Heine and Kuteva (2005). Languages may belong to one of three classes: They use two distinct markers to encode comitative and instrumental separately (= asyncretic); they do not distinguish formally between the two categories (= syncretic); or they employ two markers, of which one exclusively encodes either comitative or instrumental and the other one covers both functions (= mixed). Europe stands out from the rest of the world, as it is the only continent where the syncretic type is statistically the strongest (everywhere else the asyncretic type dominates by far). Thirty-five of the European sample languages (= 69%) are syncretic, 11 are asyncretic, and five are mixed. **Figure 4** suggests that the types are not randomly distributed over the continent but that there is a clear areal pattern.

The territory occupied by syncretic languages is divided into two subareas. The larger one covers most of the European north, west, and south. The smaller one is located in the extreme east; thus, the major homestead of the asyncretic type finds itself sandwiched between syncretic languages. What strikes the eye most is the fact that representatives of the mixed type never cluster anywhere. Where they occur they are situated on the margins of the syncretic territory, often trapped between syncretic and asyncretic languages. This is indicative of a transition from one major type to the other. Not surprisingly, the isoglosses do not respect genetic principles. The Uralic macrophylum contains syncretic, asyncretic, and mixed languages. Slavic, Celtic, Baltic, and Germanic languages are distributed over two different types. These differences in class membership are not random but rather are areally motivated. Those languages that fail to behave like their next of kin converge with their next-door neighbors, either fully or partially.

Diachronically, the properties of the syncretic type seem to be innovations, at least in a number of the languages involved. It can be shown that the feature spread from Germanic to the Baltic languages and Estonian in the early days of their documented history and triggered their typological change from asyncretic to syncretic via mixed – which is the stage reached by contemporary Lithuanian. Germanic and Italo-Romance can also be held responsible for the present

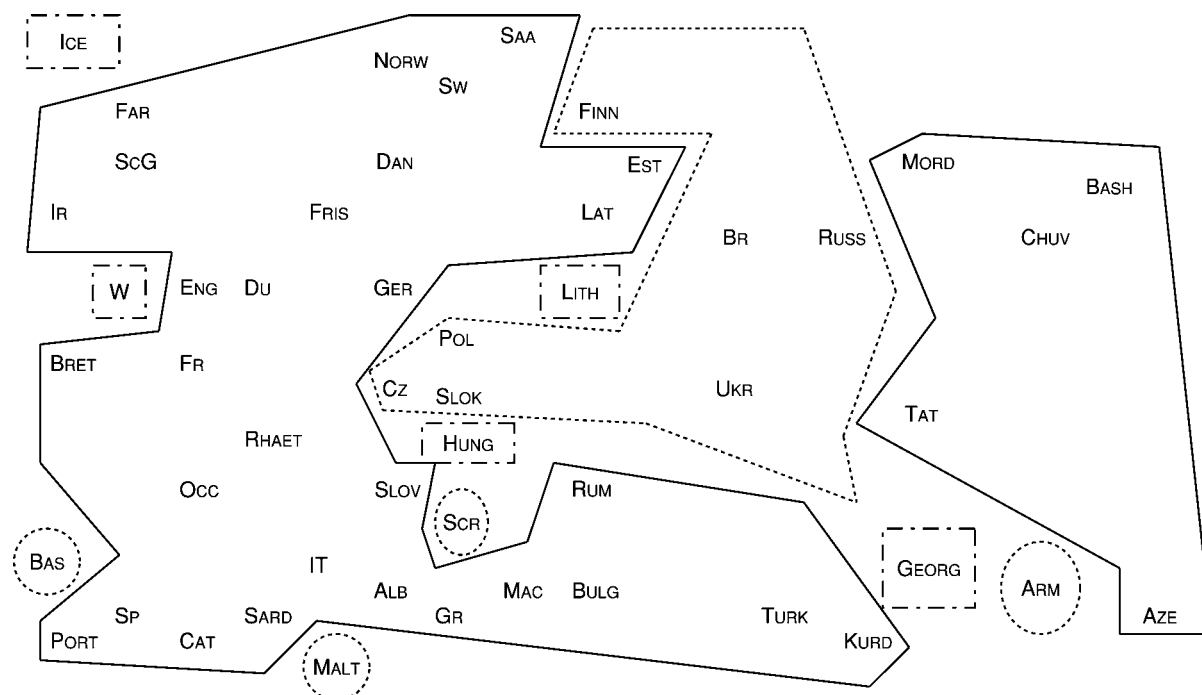


Figure 4 Comitative-instrumental syncretism [solid line = syncretic; dotted line = asyncretic; dashed/dotted line = mixed].

syncretic status of Slovenian, whereas Macedonian and Bulgarian conform to the general Balkanic picture that favors the syncretic type. A similar contact-based explanation is perhaps also possible for the mixed status of Hungarian (supported by German and Rumanian influence). In the east, syncretic properties have a genetic foundation, as all members of the Altaic phylum display the feature. Mordvin is again likely to have been influenced by the neighboring Turkic languages. Asyncretic languages are thus on the retreat.

Syncretism of comitative and instrumental is indeed a trait shared not only by the two members of the Charlemagne Sprachbund but also by those languages that belong to the more central 'layers' of SAE, including a variety of languages that do not partake regularly in isoglosses of the SAE languages.

Cardinal-based derivation of ordinals In Stolz and Veselinova (2005), the relationship between ordinal numerals and cardinal numerals in terms of derivation is studied in a crosslinguistic perspective. There are various types, the major ones being:

1. no formal distinction (= unattested in Europe)
2. ordinals are regularly derived from cardinals, as in Tatar *bër* 'one' → *bër-ěncě* 'first,' *ikě* 'two' → *ikě-ncě* 'second,' etc.
3. FIRST is not based on ONE, all other ordinals follow pattern (b), as in Georgian *erti* 'one' vs. *p'irveli* 'first' as opposed to *ori* 'two' → *me-or-e* 'second,' *sami* 'three' → *me-sam-e* 'third,' etc.

4. FIRST and SECOND are not derived from ONE and TWO, respectively, all other ordinals follow pattern (b), cf. Swedish *en* 'one' vs. *första* 'first,' *två* 'two' vs. *andra* 'second' as opposed to *tre* 'three' → *tre-dja* 'third,' *fyra* 'four' → *fjär-de* 'fourth,' etc.

On the basis of the data in Stolz (2001), Haspelmath (2001) considers the suppletive second ordinal, i.e., type (d) a potential SAE feature. Figure 5 reveals that this type is indeed the majority solution for the sample languages: 36 languages belong to type (d), 10 to type (c), and the remaining five are type (b) languages.

Type (d) occupies the large middle section on the map stretching from Saami in the north via Czech down to Sardinian in the south. The easternmost representative of this type is Mordvin, which, for once, does not show affinities to its Turkic neighbors. Irish marks the westernmost outpost of type (d) languages. The only region of Europe from which type (d) is completely missing is the southeast, where type (b) and type (c) dominate. Type (b) has a strong genetic foundation, as it is attested in all Altaic languages of the sample. The distribution of type (c) is less coherent because there are isolates and small subareas on the margins of the territories of the other two types. The largest of these comprises three West Germanic languages and Scots-Gaelic. Owing to the fact that other Celtic and Germanic languages are bona fide (d) type languages, it is clear that the isoglosses follow the familiar pattern or running

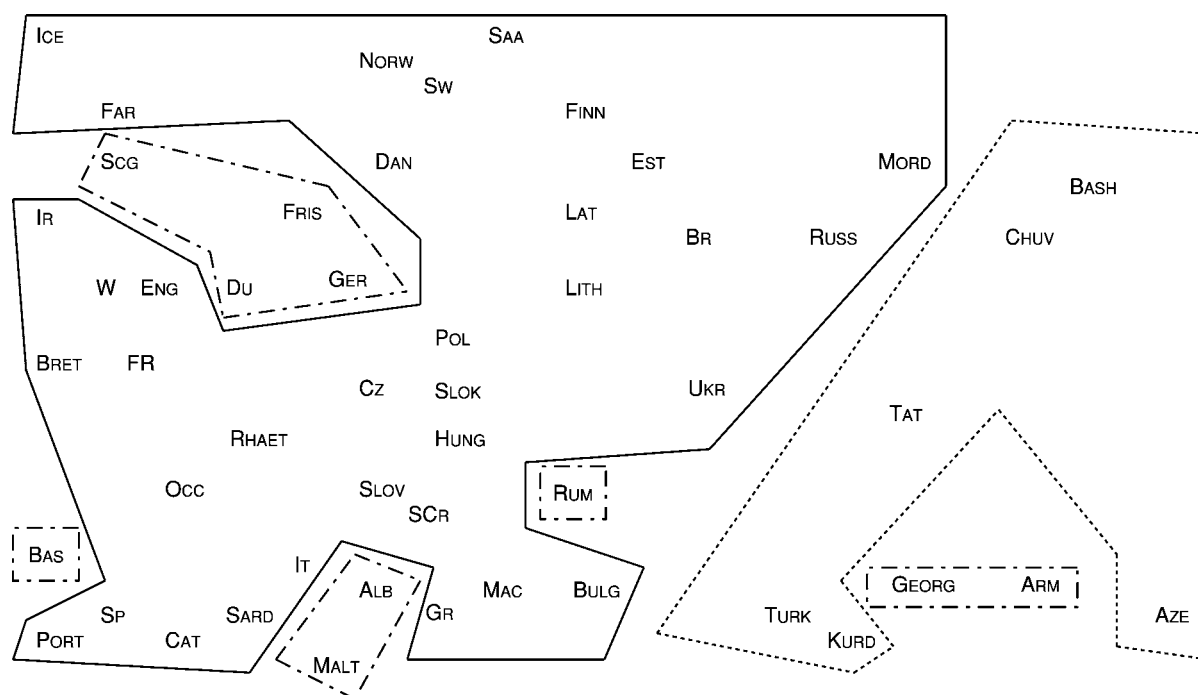


Figure 5 Suppletive ordinals [dotted line = type (b); dashed/dotted line = type (c); solid line = type (d)].

crisscross over genetic boundaries. Rumanian is the only representative of (c) type in the otherwise predominantly (d) type-oriented Romance phylum. Indo-European and non-Indo-European languages may form subareas together, as is the case for the (c) type languages Albanian and Maltese in the south and Armenian and Georgian in the southeast. The entire Uralic macrophylum partakes in the (d) type isogloss, together with the bulk of the Indo-European languages.

A word of caution is in order: There are areas of transition between the major and the minor types. Depending upon which variety of Frisian and Kurdish one chooses, the isoglosses may change direction: West Frisian follows Dutch and German (= [c] type), whereas North Frisian goes along with neighboring Danish (= [d] type). Zazaki Kurdish (Dimli) behaves like Georgian and Armenian (= [c] type), whereas other varieties of Kurdish join Turkish in the (d) type. In addition, French, with its contrast of *second* vs. *deuxième* 'second,' combines properties of both (c) type and (d) type and thus shares features not only with the bulk of typical SAE-languages that tend to belong to the (d) type but also with its partner in the *Charlemagne Sprachbund*, the (c) type language German. Turkish, Azeri, and Kurdish also allow for two allmorphs of FIRST, one of that is suppletive (Turkish *bir* 'one' → *biri-nci* 'first' vs. *ilk* 'first'). These cases add to the areality of the phenomena under scrutiny, as they clearly

demonstrate that neighborhood relations are crucial. The diachrony of the geolinguistic distribution patterns of the suppletive second ordinal still needs to be investigated more thoroughly. The most one can say for now is that German, Dutch, and Rumanian seem to have lost their erstwhile (d) type properties in the course of their history (for Rumanian, (d) type features may be found in certain stylistically marked registers).

Total reduplication Haspelmath (2001: 1503) observes that reduplication is practically unknown in contemporary European languages. This observation is correct as far as partial reduplication goes. The picture changes dramatically when we look at total reduplication, as studied in Stolz (2003). Total reduplication is firmly established as a morphological means in 25 of the sample languages, such as Maltese *ġew tnejn tnejn* {come.PERF.3Pl} {two} {two} 'they came in pairs of two,' whereas the other 26 languages do not employ total reduplication in any systematic way. For a number of languages, the situation is difficult to assess. **Figure 6** identifies the areal hotbeds of these types.

In Ibero-Romance languages, total reduplication is largely discouraged by normative grammar, although some patterns recur in the written register. This suggests that we are dealing again with an areal phenomenon. Those languages that disfavor total reduplication are located in the center of the map,

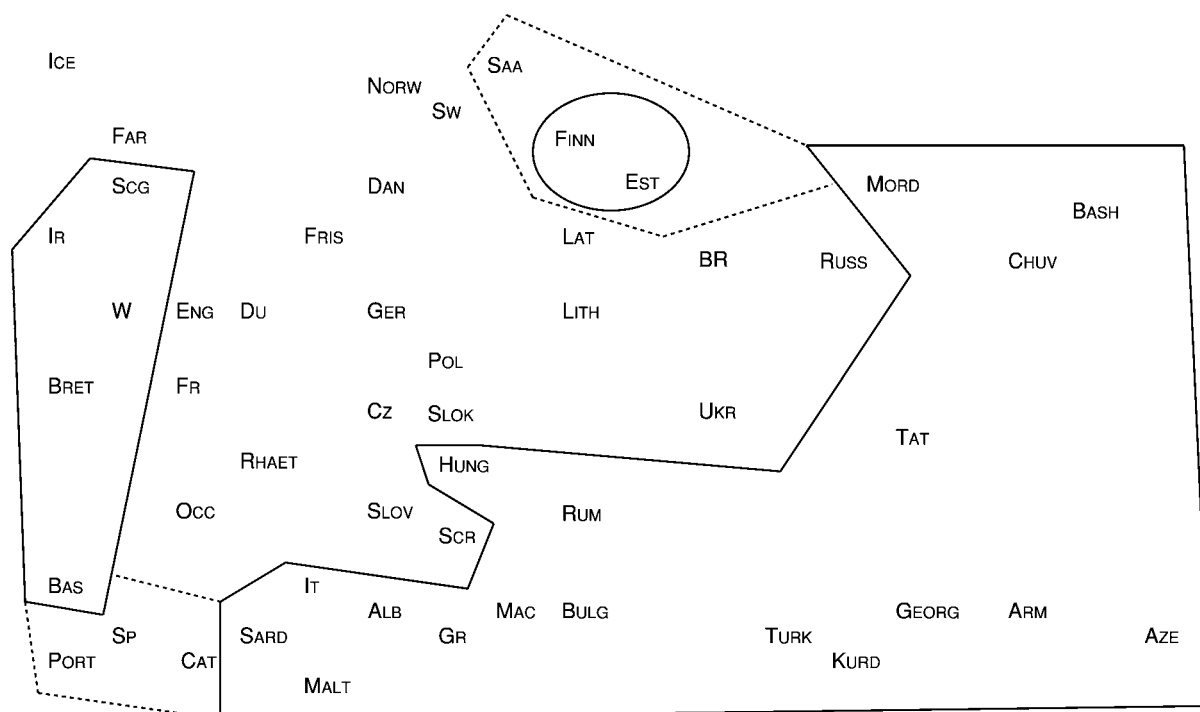


Figure 6 Total reduplication [solid line = reduplicating languages; dotted line = controversial cases].

almost completely surrounded by languages that have a predilection for total reduplication. Total reduplication is especially strong in the south and east as well as in the western rim of the continent.

As to genetic affiliation, a high percentage of languages where total reduplication is systematically employed belong to non-Indo-European macrophyla (12 out of 25 = 48% of the type). No non-Indo-European language is a bona fide case of a non-reduplicating language. Romance and Slavic phyla are divided into two, each with a substantial minority of the languages allowing for total reduplication. The entire Celtic phylum is reduplicating, whereas total reduplication is foreign to the Baltic and the Germanic phyla. Irrespective of these genetic preferences, **Figure 6** shows that neighborhood relations are again decisive: Those Romance and Slavic languages that have total reduplication share this feature with their unrelated or only distantly related neighbors. In terms of the SAE *Sprachbund*, the southern half of it is cut off from the rest by this isogloss because Italian, Sardinian, and Albanian are reduplicating but French, German, and Dutch are not.

The isogloss of total reduplication continues far beyond the limits of Europe into Siberia, the Middle East, India and South Asia, and Africa. Those languages that do not participate in this isogloss, therefore, represent the marked case. As yet, nothing definitive can be said about the history of the

isoglosses. There is evidence for diffusion of total reduplication from the Levant westward in the Mediterranean. At the same time, avoidance of total reduplication seems to spread from central Europe to the southwest. The latter processes may be supported by normative grammarians, which leaves open the question of whether total reduplication is employed by native speakers in actual discourse.

Quintessence

Summing up what **Figures 1 to 6** tell us, we observe that the distribution of the phenomena is not identical for any two isoglosses – in other words, isoglosses behave like individuals. This is of course the expected outcome. The individuality of the isoglosses notwithstanding, it is nevertheless possible to map the isoglosses onto each other to produce isopleths. **Figure 7** depicts the clustering of isoglosses from the perspective of Russian as a kind of check for the SAE-centered approach of earlier contributions of the areal linguistics of Europe. Note that every sample language shares at least one feature with Russian.

Superficially, this cluster map suggests that, with the necessary changes, it is also possible to paint a picture of the geolinguistics of Europe with a language in the center of attention that is not a prime candidate for the status of SAE language. However, there is a relatively strong genetic component that

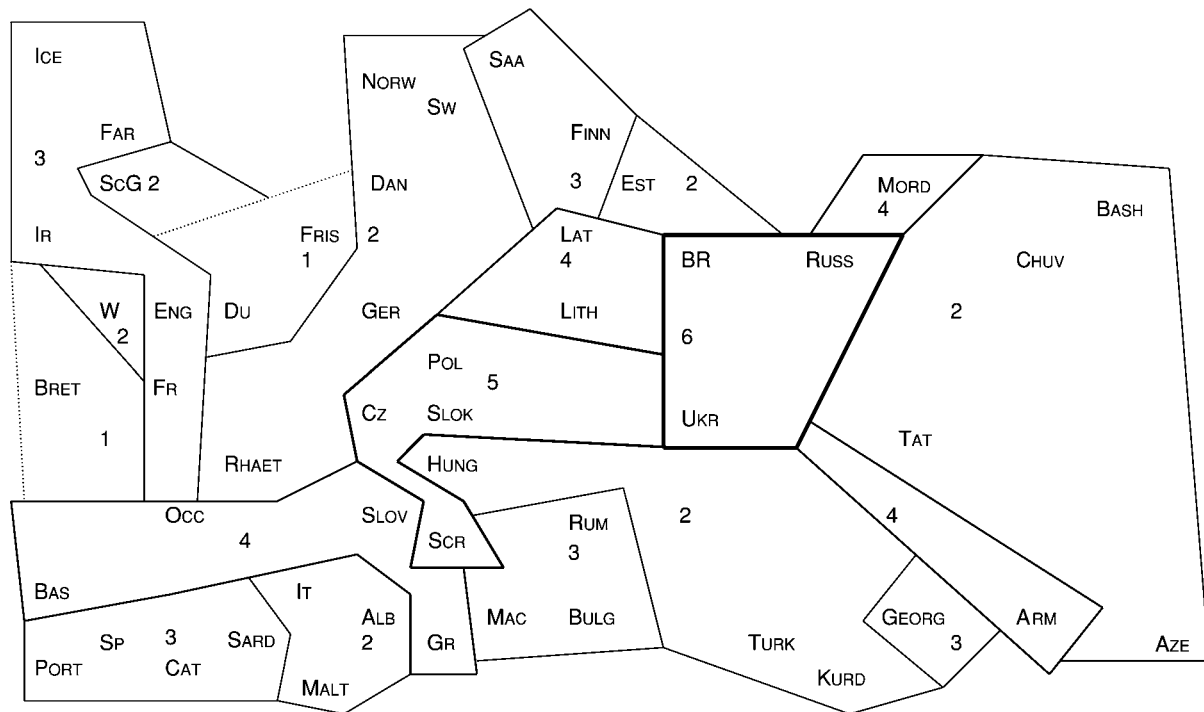


Figure 7 Cluster map.

determines similarity relations. Russian and its east Slavic sisters and neighbors, Bielarussian and Ukrainian, share all six features. The next layer, with five shared features, is also exclusively Slavic, namely the west Slavic phylum plus Serbo-Croatian. Only on the third layer does genetic diversification set in: Here we find, in addition to Slovenian, the Baltic subphylum, Armenian, Greek, all varieties of Occitan, Basque, and Mordvin. The smaller the number of shared features, the higher the genetic heterogeneity of the layer. Interestingly, the languages with the minimal number of one single shared feature, Dutch, Frisian, and Breton, are all spoken in the west at a distance from Russian. Mainland Scandinavian languages and central Europe are likewise less prone to associate with Russian in isoglosses. Nevertheless, there seems to be a somewhat stronger westward orientation of the assumed area based on Russian, as the members of the Turkic phylum in the east display unexpectedly low degrees of similarity to Russian. Discounting the fact that the very first layers (with five or six shared features) on the cluster map are clearly grounded in the Slavic phylum, it is nevertheless possible to postulate an areal structure with decreasing numbers of shared features from center to periphery. Thus, this still very superficial look at Europe from the eastern vantage point supports the idea that there is a counterpoint to SAE. This view is in line with Lewy's ([1942] 1964) observation that there

is an east-west asymmetry in Europe. As Haspelmath (2001: 1505) observes, the coexisting areas collide and overlap (most probably on the Balkans) and thus contribute to the geolinguistic diversity of Europe.

Europe as a Contact-Superposition Zone

From the above we learn that one and the same language may be part of residual and expansive areas, depending on the feature under review. What the data seem to suggest is the frequent diffusion of features originating on the (south)western mainland of the continent to the north and east. With a view to determining the balance or imbalance of SAE-based innovations and non-SAE-based ones, a much more detailed in-depth study of the geolinguistics of Europe is called for.

In the absence of such a study, one can tentatively conclude that:

- Europe is not a homogeneous linguistic area on a par with, say, the Balkan *Sprachbund*.
- The areal linguistics of Europe cannot be reduced to the identification of SAE features because:
- many languages of the continent are never or hardly ever included in the relevant isoglosses;
- many important features have not yet been checked for their areal distribution within the confines of Europe;

- the languages that fail to qualify as more central SAE languages (including those that do not share a single of the features characteristic of SAE) are likely candidates for at least one non-SAE linguistic area.
- The feature/isogloss-based approach to areal linguistics is methodologically superior to others, as it is based on proper linguistic criteria without the necessity of recourse to nonlinguistic ones.
- The feature/isogloss-orientation also gives us the opportunity to look beyond the geographic boundaries of Europe in order to establish whether the languages involved form part of a much larger macro area (Kuteva, 1998).

Given that isoglosses overlap and relatively seldom come in bundles, and given too that features may originate in different places, it is probably more accurate to speak of Europe as a contact-superposition zone (Koptjevskaja-Tamm and Wälchli, 2001) in lieu of using the suggestive label *Sprachbund* or even the somewhat less restrictive term 'linguistic area.' The present distribution of features on the European map is the product of a variety of processes that happened at different epochs. Not all of the phenomena can be attributed to the time of the Great Migrations or to any one major historical event of the distant past (Haspelmath, 1998; Haspelmath, 2001: 1506–1507). Some innovations spread at a higher speed than others, some are attractive and are copied in language contact, others fail to meet this criterion and thus never extend beyond a certain region. One can therefore agree with Haspelmath (2001: 1507) when he concludes that the distribution of different features is 'due to different historical circumstances, and the correct picture is likely to be much more complicated than we can imagine at the moment.'

With a view to coming closer to the correct picture, future investigations on the basis of a much larger sample of languages will have to integrate both evidence from substandard varieties and diachronic data.

Bibliography

- Andersen H (ed.) (1985). *Sandhi phenomena in the languages of Europe*. Berlin: Mouton de Gruyter.
- Auer P (2004). 'Sprache, Grenze, Raum.' *Zeitschrift für Sprachwissenschaft* 23(2), 149–179.
- Banfi E & Grandi N (2003). *Lingue d'Europa: elementi di storia e di tipologia linguistica*. Roma: Carocci.
- Bechert J ([1981] 1998). 'Notiz über eine Möglichkeit, die historisch-vergleichende Sprachwissenschaft zu vervollständigen, oder: Lesefrüchte zur Verbesserung Mittel-Europas und anderer Weltgegenden.' In Boeder W *et al.*

- (eds.) *In memoriam Johannes Bechert. Sprache in Raum und Zeit. Band 1: Johannes Bechert: Kleine Schriften*. Tübingen: Narr. 13–16.
- Becker H (1948). *Der Sprachbund*. Leipzig: Humboldt.
- Bernini G & Ramat P (1992). *La frase negativa nelle lingue d'Europa*. Bologna: Il Mulino.
- Bossong G & Comrie B (1998–2003). *EUROTYP publications 20/1–8*. Berlin: Mouton de Gruyter.
- Chambers J K & Trudgill P (1980). *Dialectology*. Cambridge: Cambridge University Press.
- Cristofaro S (2000). 'Linguistic areas, typology and historical linguistics: an overview with particular respect to Mediterranean languages.' In Cristofaro S & Putzu I (eds.) *Languages in the Mediterranean area: typology and convergence. Il progetto MEDTYP: studio dell'area linguistica mediterranea*. Milano: Franco Angeli. 65–81.
- Dahl Ö (1990). 'Standard Average European as an exotic language.' In Bechert J, Bernini G & Buridant C (eds.) *Toward a typology of European languages*. Berlin: Mouton de Gruyter. 3–8.
- Décsey G ([1973] 2000). *The linguistic identity of Europe* (2 vols). Bloomington: Eurolingua.
- Haarmann H (1976a). *Grundzüge der Sprachtypologie. Methodik, Empirie und Systematik der Sprachen Europas*. Stuttgart: Kohlhammer.
- Haarmann H (1976b). *Aspekte der Arealtypologie. Die Problematik der europäischen Sprachbünde*. Tübingen: Narr.
- Haarmann H (1977). *Prinzipielle Probleme des multilateralen Sprachvergleichs*. Tübingen: Narr.
- Haarmann H (1986). 'Allgemeine Strukturen europäischer Standardsprachentwicklung.' *Sociolinguistica* 2, 10–51.
- Haspelmath M (1998). 'How young is Standard Average European?' *Language Science* 20, 271–287.
- Haspelmath M (2001). 'The European linguistic area: Standard Average European.' In Haspelmath M, König E, Oesterreicher W & Raible W (eds.) *Language typology and language universals: an international handbook*, vol. 2. Berlin: Mouton de Gruyter. 1492–1510.
- Heine B & Kuteva T (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Hinrichs U (ed.) (2004). *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*. Wiesbaden: Harrassowitz.
- Jakobson R ([1931] 1971). 'Karakteristike evrazijskogo jazykovogo sojuza.' In Jakobson R (ed.) *Selected writings*, vol. 1. The Hague: Mouton. 144–201.
- König E & Haspelmath M (1999). 'Der europäische Sprachbund.' In Reiter N (ed.) *Eurolinguistik. Ein Schritt in die Zukunft*. Wiesbaden: Harrassowitz. 111–127.
- Koptjevskaja-Tamm M & Wälchli B (2001). 'The Circum-Baltic languages: an areal-typological approach.' In Dahl Ö & Koptjevskaja-Tamm M (eds.) *Circum-Baltic languages*, vol. 2. Amsterdam: Benjamins. 615–750.
- Kortmann B (ed.) (2004). *Dialectology meets typology: dialect grammar from a crosslinguistic perspective*. Berlin: Mouton de Gruyter.

- Kuteva T (1998). 'Large linguistic areas in grammaticalization: auxiliation in Europe.' *Language Science* 20, 289–311.
- Lewy E ([1942] 1964). *Der Bau der europäischen Sprachen*. Tübingen: Niemeyer.
- Matras Y, McMahon A & Vincent N (eds.) (2005). *Linguistic areas*. Houndsmill: Palgrave Macmillan.
- Mayerthaler W, Fliedl G & Winkler C (1993). *Infinitivprominenz in europäischen Sprachen. Teil 1: Die Romania (samt Baskisch)*. Tübingen: Narr.
- Nocentini A (2002). *L'Europa linguistica: profilo storico e tipologico*. Firenze: Le Monnier.
- Pott A F (1868). *Die Sprachverschiedenheit in Europa an den Zahlwörtern nachgewiesen sowie quinäre und vigesimale Zählmethode*. Halle: Waisenhaus.
- Pott A F (1887). *Zur Litteratur der Sprachenkunde Europas*. Leipzig: Teubner.
- Ramat P & Stolz T (eds.) (2002). *Mediterranean languages*. Bochum: Brockmeyer.
- Sandfeld K (1930). *Linguistique balkanique: problèmes et résultats*. Paris: Champion.
- Sarhimaa A (1991). 'Karelian Sprachbund? Theoretical basis of the study of Russian/Baltic–Finnic contacts.' *Finnisch-Ugrische Forschungen* 50, 209–219.
- Schleicher A (1850). *Die Sprachen Europas in systematischer Übersicht: linguistische Untersuchungen*. Bonn: König.
- Siewierska A (1998). *Constituent order in the languages of Europe*. Berlin: Mouton de Gruyter.
- Skalička V (1968). 'Zum Problem des Donausprachbundes.' *Ural-Altaische Jahrbücher* 40, 3–9.
- Stadnik E (2002). *Die Palatalisierung in den Sprachen Europas und Asiens. Eine arealtypologische Untersuchung*. Tübingen: Narr.
- Stolz T (2001). 'Ordinalia – linguistisches Neuland: Ein Typologienblick auf die Beziehung zwischen Kardinalia und Ordinalia und die Sonderstellung von eins und erster.' In Igla B & Stolz T (eds.) *Was ich noch sagen wollte ... Festschrift für Norbert Boretzky zu seinem 65. Geburtstag*. Berlin: Akademie. 507–530.
- Stolz T (2002). 'Review of Gyula Décsy *The linguistic identity of Europe*. 2 vols. Bloomington: Eurolingua.' *Anthropological Linguistics* 44, 308–320.
- Stolz T (2003). 'A new Mediterraneanism: word iteration in an areal perspective. A pilot-study.' *Mediterranean Language Review* 15, 1–47.
- Stolz T, Stroh C & Urdze A (2003). 'Solidaritäten.' *Lingua Posnaniensis* 45, 69–92.
- Stolz, T. & Veselinova, L. (2005). 'Ordinal numerals.' In Comrie, B. et al. (eds.) *World atlas of language structures*. Oxford: Oxford University Press.
- Ternes E (1998). 'Lauttypologie der Sprachen Europas.' In Boeder W et al. (eds.) *In memoriam Johannes Bechert. Sprache in Raum und Zeit. Band 2: Beiträge zur empirischen Sprachwissenschaft*. Tübingen: Narr. 139–152.
- Ureland P S (2004). 'Eurolinguistik und Europäistik als Fächer an den Universitäten.' In Hinrichs U (ed.) *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*. Wiesbaden: Harrassowitz. 1–15.
- Van der Auwera J (1998). 'Conclusion.' In van der Auwera J (ed.) *Adverbial constructions in the languages of Europe*. Berlin: Mouton de Gruyter. 813–836.
- Wagner H (1959). *Das Verbum in den Sprachen der britischen Inseln: ein Beitrag zur geographischen Typologie des Verbums*. Tübingen: Niemeyer.
- Wagner H (1964). 'Nordeuropäische Lautgeographie.' *Zeitschrift für Celtische Philologie* 29, 225–298.
- Whorf B L ([1941] 1956). 'The relation of habitual thought and behavior to language.' In Carroll J B (ed.) *Language, thought and reality: selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press. 134–159.
- Wintschalek W (1993). *Die Areallinguistik am Beispiel syntaktischer Übereinstimmungen im Wolga-Kama-Areal*. Wiesbaden: Harrassowitz.

Evenki

I Nedjalkov, Institute of Linguistic Research,
St Petersburg, Russia

© 2006 Elsevier Ltd. All rights reserved.

Evenki belongs to the Tungusic family, widely considered to form a branch of the Altaic languages. Tungusic family comprises three subgroups, the Northern (or Siberian), the Southern (or Amur), and the Manchu. (There is another classification in which two subgroups are distinguished, the Northern group and the Southern group including Manchu.) The number of native speakers of Evenki in Russia does not exceed 30 000. Evenks live on vast territories in

Siberia, far east of Russia, and in the north of China and Mongolia (where they are called Oroqs or Orochons; about 12 000). The Evenki Autonomous Region comprises about 768 000 square kilometres; its population is about 30 000 and only about 5000 of them are Evenks. If we sum up all the territories of Siberia and the far east of Russia inhabited by the Evenks, the total will equal the territory of at least one-third of Russia. There is hardly another people in the world as small as the Evenks that is aboriginal to such a vast area, as they were a nomadic people.

Evenki is also remarkable for its number of dialects and subdialects, about 50 in all. They are subdivided into three groups, Northern, Southern, and Eastern.

The Northern dialects of Evenki are spoken in the northern part of the Krasnojarsk and Irkutsk regions, and the Southern dialects around Lake Baikal and in Buryatia. The Eastern dialects are spoken in the Republic of Sakha-Yakutia, in the Amur and Khabarovsk regions and on the Island of Sakhalin. The Evenki language acquired its writing system in the early 1920s. It was based on the Latin alphabet but later (in the early 1930s) it was replaced by the Cyrillic alphabet. Nowadays, books and newspapers are published in Evenki.

General Characteristics: Sentence Structure and Morphology

Evenki is an agglutinating (suffixal) language with no prefixes. It has nonrigid SOV word order, rich verbal morphology, and predominantly participial and converbal syntax. Adjectives, demonstrative, and possessive pronouns, and numerals always precede the head noun.

Case, Number, and Possessivity

The Evenki noun has 13 cases (plus the comitative, which can also be viewed as a special noncase form). Nominative: zero marker (marks the subject); Accusative 1: *-va* (marks the definite direct object); Accusative 2: *-a/-ya* (marks the indefinite direct object; can also express the partitive meaning and, with markers of personal or reflexive possession, the benefactive meaning of the direct object); Dative: *-du/-tu* (marks locative and temporal adverbials and also addressee and beneficiary); Allative 1: *-dula/-tula/-la*; Allative 2: *-tki*; Allative-Prolative: *-kli*; Locative-Allative: *-kla*; Instrumental: *-t/-di*; Ablative 1: *-duk/-tuk*; Ablative 2: *-git*; Prolative: *-duli/-li*; Comitative: *-mun*.

Accusative 2 (traditionally termed indefinite accusative) is used either for indefinite nonreferential objects or for a partitive meaning, and it generally occurs either with the future tense or with the imperative; e.g.,:

- (1) *Ukumni-ye min-du bu:-kel*
milk-ACC2 I-DAT give-IMPERATIVE:2SG
'Give me some milk'

Accusative 2 with the markers of personal possession codes object-oriented benefactive forms, while the same case with reflexive-possessive markers codes subject-oriented benefactive forms, e.g.,: *dyav-ya-v* 'a boat for me', *dyav-ya-s* 'a boat for you' (SG); *dyav-ya-vi* 'a boat for oneself (myself/yourself/himself/herself)'; *dyav-ya-var* 'a boat for ourselves/your-selves/themselves.' The plural marker is the suffix *-l* on the absolute majority of nominal stems; on nouns with the stem-final *-n*, the plural marker is *-r*, which ousts *-n*, e.g., *oron* 'reindeer.SG' → *oro -r* 'reindeer-PL'. The personal-possessive affixes are: *dyu-v* 'my house', *dyu-vun* 'our (EXCL) house'; *dyu-t* 'our (INCL) house', *dyu-s* 'your (SG) house', *dyu-sun* 'your (PL) house', *dyu-n* 'his/her house', *dyu-tyn* 'their house.' Reflexive-possessive affixes are *-vi/-mi* for a singular possessor: depending on the person of the subject it may mean 'my', 'your (SG)', 'his/her'; and *-var/-mar* for a plural possessor: depending on the person of the subject it may mean 'our', 'your (PL)', or 'their'. Morphemic ordering in nouns is the following: noun stem - plural - case - possession, e.g.,: *dyu-l-dula-tyn* 'to their houses/tents'.

Tense/Aspect System and Agreement

There are eight tenses; the markers are: *-ra*: non-future tense; *-dyara*: present tense; *-cha*: past tense; *-dyacha*: imperfect; *-ngki*: iterative past; *-dya*: future 1; *-dyanga*: future 2; *-dyalla*: future 3. There are about 10 aspectual markers: imperfective *-dya*; inchoative *-l*; semelfactive *-sin/-sn/-s*; distributive *-kta*; durative *-t/-chi*; habitual *-ngna*; iterative *-van/-vat*; resultative *-cha*; quick action *-malcha*.

There are two types of agreement. The first one coincides with the personal possession nominal markers and is used with tense forms that go back to (and coincide with) participles. The second type is the system of verbal agreement markers proper. Agreement markers of the finite verb forms are the following (see Table 1).

Non-finite Verb Forms

There are about 15 converbs and 10 participles. The most common participles are the habitual

Table 1 Agreement markers

| | Nominal type | | Verbal type | |
|--------|--------------|----------------|-------------|--------------|
| | SG | PL | SG | PL |
| 1st p. | <i>-v</i> | <i>-vun/-t</i> | <i>-m</i> | <i>-v/-p</i> |
| 2nd p. | <i>-s</i> | <i>-sun</i> | <i>-nni</i> | <i>-s</i> |
| 3rd p. | <i>-n</i> | <i>-tyn</i> | <i>-n</i> | <i>-φ</i> |

(marker-*wki*), the simultaneous (-*dyari*), the anterior (-*cha*), the posterior (-*dyanga*). The most common converbs are those of anteriority (marked by -*ksa/kanim*) and temporal-conditional converbs in -*mi* (same-subject) and -*raki* (different-subject).

Voice System, Means of Valency Change, and Their Combinability

There are four productive means of changing the valency and/or the number of participants, traditionally regarded as voices: causative (marker -*vkan*), passive/decausative (-*v/-p/-mu*), reciprocal (-*mat/-mach*), and sociative. The sociative marker -*ldy* does not change syntactic valency of the base verb, but it changes the number of participants. Reflexivity is expressed pronominally. In this respect, Evenki, like other Tungusic languages, differs from the neighboring Turkic languages and reveals similarity to the neighboring Mongolian languages.

Causative derivation increases the valency of the base verb by one. Causatives are freely derived from all pure transitive and intransitive stems (i.e., stems with no other voice/valency suffixes) and also from sociatives and a few reciprocals, e.g.: *iche*- 'to see sth' → *iche-vken*- 'to show sth to sb' → *iche-vken-met*- 'to show sth to each other'; *archa*- 'to meet sb' → *archa-mat*- 'to meet each other' → *archa-machi-vkan*- 'to cause/let/make sb (to) meet each other'.

Sociative derivation is possible from all pure verb stems and also from causatives, but not from reciprocals and passives:

- (2) *Asa-l dyu-va*
 woman-PL house-ACC
iche-vke-ldy-re-0
 see-CAUS-SOC-NFUT-3PL
 'The women showed the house [to someone else] together'

Passive derivation (suffix -*v*) is possible from all transitive stems including causatives (thus resulting in either personal or impersonal passive constructions), almost all intransitives (resulting in impersonal passive constructions only) and seven intransitive 'weather' verbs (resulting in personal adversative passives). The marker -*v* homonymous with the passive suffix can function as a nonproductive causative suffix. Passives are not derived from reciprocals and sociative stems.

- (3a) *Asi dyu-va o:-ra-n*
 woman house-ACC make-NFUT-3SG
 'The woman put up a tent'

- (3b) *Dyu o:-v-ra-n*
 tent make-PASS-NFUT-3SG
 'The tent was put up'/'The house was built'

Reciprocal derivation is impossible from passive and sociative stems (the suffix -*ldy-met* is not a reciprocal derivation from sociative but a complex reciprocal suffix). It is possible from causative stems:

- (4a) *Asa-l dyu-l-var*
 woman-PL house-PL-their
iche-vken-met-te-0
 see-CAUS-REC-NFUT-3PL
 'The women showed their houses to each other'
- (4b) *Nungartyn eme-vken-met-chere-0*
 they come-CAUS-REC-PRES-3PL
 'They cause each other to come'

Anticausatives from verbs denoting destruction or change of state are formed by means of the suffix -*rga/-rge*. From a small group of verbs, anticausatives are formed by means of the suffix -*v/-p/-mu*, and it also functions as a passive marker (note that this very suffix is also used to derive causatives from about 50 intransitive and 20 transitive verbs):

- (5a) *kapu*- 'to break' (vt) → *kapu-rga* 'to break' (vi)
ety 'to tear' (vt) → *ety-rge* 'to tear' (vi)
- (5b) *das*- 'to close' (vt) → *dasi-v*- 'to close' (vi)
sukcha 'to break' (vt) → *sukcha-v*- 'to break' (vi)

Negation, Modality Markers, and Morphemic Ordering

There are two major ways of expressing negation in Evenki: (a) by means of the conjugated negative auxiliary verb *e* 'not to ...', and (b) with the negative noun *achin* 'no'.

The modality markers include: -*mu* ('want'), -*ssa* ('try'), and -*na* ('go'), e.g., Nungan homoty-va (ACC1) va: -*na-ssa-mu-dyere-n* lit. 'He wants to try to go and kill the bear'.

Morphemic ordering in verbs is the following: verb stem - causative - sociative - reciprocal - aspect - passive - modality - evaluation - aktionsarten - tense or non-indicative moods or nonfinite markers - subject agreement (in person/number).

Bibliography

- Kolesnikova V D (1966). *Sintaksis evenkijskogo jazyka* [Syntax of the Evenki language]. Moskva-Leningrad.
 Konstantinova O A (1964). *Evenkijskij jazyk* [The Evenki language]. Moskva-Leningrad: Nauka.

Lebedeva E P, Konstantinova O A & Monakhova I V (1985). *Evenkijskij jazyk* [The Evenki language]. Leningrad: Prosveščenie.
Nedjalkov I V (1997). *Evenki*. London: Routledge.

Vasilevič G M (1958). *Evenkijsko-russkij slovar'* [Evenki-Russian dictionary]. Moskva.
Vasilevič G M (1966). *Istoričeskij folklor evenkov* [Historical folklore of Evenks]. Moskva-Leningrad: Nauka.

Ewe

J Essegbey, University of Florida,
Gainesville, FL, USA

© 2006 Elsevier Ltd. All rights reserved.

Also written *Eve*, and pronounced /əβə/, Ewe designates a dialect cluster that is spoken mainly in South-eastern Ghana, but also in the southern part of Togo and across the Togo-Benin border. Its dialects include Kpele and Notsie, spoken in Togo, Waci, spoken in Benin, and Aɲɔ, Tɔɲu, Ho, Kpedze, Anfoe, and Kpandu, all spoken in Ghana. The dialects in Ghana are grouped into Coastal (e.g., Aɲɔ, Tɔɲu), Central (e.g., Ho, Kpedze), and Northern (e.g., Anfoe, Kpando). The Central and Northern dialects are also grouped together as Inland dialects. Ewe is part of the Gbe language cluster (cf. Capo, 1991), which belongs to the Kwa family.

History and Sociolinguistics

The origin of the Gbe-speaking people has been traced to Ketu, a Yoruba town in the present-day Benin. From there they moved southward, with some founding a settlement at Tado, while others settled at Adele and others went to Notsie. The Ewes are among those who settled in Notsie. They say that their forefathers fled the tyranny of a ruler of Notsie called Agorkoli, and dispersed to their present-day locations.

Ewe is spoken by approximately 3–5 million speakers. It is taught as a subject from elementary school to the university level and is one of the seven national languages in the media in Ghana, and one of two in Togo. A clear dialectal difference can be established between the Coastal and Inland dialects, e.g., 'ash' is *afi* in the Coastal dialects, and *dzowɔ* in the Inland. The former also have a habitual suffix *-na*, which is represented in the other dialects as *-V*, the resultant realization being determined by the preceding vowel. Hence *zɔna* 'walks' is *zɔɔ* in the Inland while *fena* 'plays' is *fee*.

In addition, the initiator of greetings at the Coast exhausts all his or her questions before the interlocutor begins. By contrast, in the inland, both speakers take turns in asking how-are-you questions.

Left-hand use is prohibited in social interaction, giving rise to interesting modes of pointing.

Phonology

Ewe has 29 consonants and seven oral vowels with nasalized counterparts. Notable among the consonants are labiovelar stops, bilabial fricatives, and a velar approximant. Both /l/ and /r/ occur in complementary distribution: /r/ occurs before laminodentals, alveolars, and palatals, while /l/ occurs elsewhere, including in word-initial position. Also, /w/ occurs before rounded vowels while /y/ occurs before unrounded vowels (*wɔ* 'do' vs. *yi* 'white'). Although most of the dialects have all the nasalized vowels, a few, including Peki, do not have /õ/. Thus instead of *lõ* 'take off fire,' they say *lɔ*.

Ewe is a tone language. Phonetically, all the dialects have high, mid, and low tones, which also combine to yield six contour tones. Phonologically, it has a high and nonhigh tone, both of which are dependent on the environment. Thus nonhigh tone is realized as mid in a root noun that has a voiceless obstruent or sonorant, and low before a voiced obstruent. High tone is realized as high before a voiceless obstruent and mid before a voiced obstruent. Aɲɔ also has a phonologically conditioned extrahigh tone, while Adangbe has an extralow tone that occurs on the utterance-final interrogative particle.

The syllable structure is mainly open, although a few words end with nasals, e.g., *kɔpam* 'sound of slap.' While it is possible to have a maximum of two syllable-initial consonants, the second has to be a liquid or approximant. The nucleus can be a vowel or a nasal.

Morphology

Ewe has been characterized as an isolating language with agglutinating features (Ameka, 1991). Thus, many words look like a concatenation of individual morphemes:

nyɔnu-vi-wo
woman-little-PL
'girls'

It has only one inflectional affix, i.e., the habitual suffix *-na*. However, it has derivational processes, which include reduplication, triplication, and compounding. It also has derivational affixes such as the agentive *la*:

nu-fia-la
[thing-teach-AG]
teacher

Syntax

Syntactically, Ewe is an SVO language, with alternative OSV order being determined by semantic and pragmatic factors such as topicalization and focusing. Within the noun phrase, modifiers follow the head noun (*qevi nyui la* 'child good the'). The plural morpheme *wo* is related to the third person plural pronoun *wo* and is not required when the NP has a numeral (*qeviwo* 'children' vs. *qevi eve* 'two children'). It is, however, obligatory after a determiner (*qevi eve ma-*(wo)* 'child two that-PL'). There is a logophoric pronoun (*ye*) that occurs in a subordinate clause introduced by *be(na)* 'that' (cf. Clements, 1979; Essegbey, 1994).

Ewe is a tenseless language. An active verb in the aorist receives past tense interpretation, (*e-qu nu* 'he ate'), while a stative or inchoative verb receives present tense interpretation (*e-ku* 'it is dead'). A potential morpheme gives rise to future interpretation. It has a serial verb construction (SVC) in which the two or more verbs in a clause share the same TMA value. Negation in the clause is a discontinuous morpheme *me ... o*: *me* precedes the first verb and *o* occurs at the

end of the clause. Ewe has obligatory complement verbs (OCVs): verbs with fully specified meaning have to take a generic-meaning complement (*qu nu* 'eat thing'), while others with less determined meanings have their meaning further specified by the complement. These are known as inherent complement verbs (ICVs), (*fu tsi* 'move_limb water = swim').

There are two types of double object constructions: Theme-Goal (*bia nya Kofi*, literally 'ask word Kofi') and Goal-Theme (*na Kofi ga* 'give Kofi'). *Na* 'give' and *fia* 'teach' occur in both constructions. Finally, Ewe has ideophones, some of which code manner of motion concepts, e.g., *dziadzia* 'energetic walking.'

Bibliography

- Ameka F K (1991). *Ewe: its grammatical constructions and illocutionary devices*. Ph.D. diss., Australian National University, Canberra.
- Capo H B C (1991). *A comparative phonology of Gbe*. Berlin and Garome: de Gruyter (Foris) and LABOGBE.
- Clements G N (1979). 'The logophoric pronoun in Ewe: its role is discourse.' *Journal of West African Languages* 10, 141–171.
- Essegbey J (1994). *Anaphoric phenomena in Ewe*. M.A. Thesis. University of Trondheim.
- Essegbey J (1999). *Inherent complement verbs revisited: towards an understanding of argument structure in Ewe*. Ph.D. dissertation. Leiden University.
- Essegbey J (2004). 'Auxiliaries in serialising languages: on COME and GO verbs in Sranan and Ewe.' *Lingua* 114, 473–494.
- Kita S & Essegbey J (2001). 'Pointing left in Ghana: how a taboo on the use of the left hand influences gestural practices.' *Gesture* 1(1), 73–95.

This page intentionally left blank

F

Fanagalo

R Mesthrie, University of Cape Town, Cape Town, South Africa

© 2006 Elsevier Ltd. All rights reserved.

Fanakalo (also spelled Fanagalo) is a southern African pidgin language that continues to be used two centuries after its inception. It is used in parts of South Africa, Zimbabwe (where it is usually known as Chilapalapa), Mozambique, Malawi, Zambia, and Namibia where it has been carried by migrant workers in the South African mines. Within South Africa it is spoken mainly in the provinces of KwaZulu-Natal and Gauteng (the mining area).

Fanakalo can be described as a 'crystallized' pidgin in terms of its fairly stable structure and circumscribed contexts of use. It is a contact language used prototypically in work situations: on farms, in the mines of the Witwatersrand that draw a multilingual workforce from all over southern Africa, in other urban labor situations, and in domestic employment (between employers and maids, cooks, gardeners). One can hear it in situations of sustained labor contacts, as well as in 'transactional' communications as in gas stations, shops, markets, and the like. In South Africa in provinces other than KwaZulu Natal and Gauteng, Fanakalo is less well known, as the rival urban lingua franca in the domain of labor is Afrikaans. In rural areas, population demographics often dictate that white farmers and their families acquire the local Bantu language, especially Tswana in the Free State, Xhosa in the eastern Cape, and Zulu in KwaZulu-Natal. In former times, Fanakalo was also used in nonlabor contexts when Europeans and/or Indians had no other means of communication with each other. Thus, it was used sporadically "by white men amongst themselves when no other means of communication are available" (Mayne, 1947: ii) and when North Indians had no other means of communication with South Indians (Mesthrie, 1989).

Fanakalo use is receding slightly as English spreads as a lingua franca among younger people, even on farms. In addition, its use is no longer officially sanctioned in the mines of post-apartheid South Africa

because of its long-standing association with cheap labor and racism. However, there are still ample situations in which it is used, including some non-labor contexts (Adendorff, 1995). Two such uses are a kind of expatriate solidarity or nostalgia for things South African expressed by some expatriates in light-hearted communications with family in South Africa or as a secret language for Zimbabwean or KwaZulu-Natal tourists abroad, especially for those who do not speak a Bantu language or Afrikaans. Fanakalo is sometimes used by Zulu speakers as a playful form of code divergence that signifies harsher relations with interlocutors than is possible using Zulu (Adendorff, 1995).

Origins

The first sustained contacts between Europeans and indigenous peoples in South Africa took place in the western Cape in the 17th century, where Afrikaans arose as a lingua franca out of the experience of colonization and slavery. When Afrikaners moved into the eastern Cape from about 1770 onward, Afrikaans was no longer a viable means of communication with the Xhosa people, and several strategies of communication arose: by signs, by simplified Xhosa, by simplified Afrikaans, or by a mixture of these methods (Mesthrie, 1998). With the arrival of the first batch of settlers from England, English was added to the frontier in 1820.

Fanakalo probably came into existence amid these diffuse communicational circumstances in the eastern Cape in the early 1800s. Mesthrie (1998: 13) gave the earliest recorded sentence in the pidgin as *Wena tandaza O Taay* 'You (must) worship God' (uttered by the missionary John Reid, Kat River 1816, who thought he was speaking Xhosa). Fanakalo does not seem to have been widespread in this period: It is but one of several communication strategies that appear in the archival and travel literature of the times, and judging from the sources it was used not very frequently.

Among the many diffuse strategies of communication on the eastern Cape frontier, the one that won

out later in the new colony of Natal (established in 1843 further north along the coast) was the Fanakalo option. The pidgin, which was initially known as 'Kitchen Kaffir' in this colony, was likely brought over by people with experience of the frontier: Afrikaners away from the British in the Cape Colony, their 'colored' servants, English adventurers, and possibly some officials. No concrete evidence of this link exists, however, and the brief accounts of Fanakalo give a picture of a pidgin being invented anew from contacts between British settlers and the Zulus who outnumbered them. Two major crystalizing events for Fanakalo took place in this period: (1) the arrival of indentured Indians in large numbers in the coastal province of Natal (starting in 1860); and (2) the discovery of diamonds and gold in the interior (starting in 1867).

Structure

Although Fanakalo has none of the inflexional/agglutinative richness of Zulu (see Cole, 1953), it is not as impoverished as one might expect of a pidgin. It has four tense markers used with verbs, which are all derived from Zulu (or Nguni languages generally): *-a* (for infinitive, imperative, and present tense verbs); *-ile* (past tense); *zo* (future); and *gate* (anterior). The first two tense markers are suffixes, whereas the future marker *zo* occurs either as a free form that precedes the verb or as [z] cliticized to the subject pronoun. The fourth tense marker, *gate* (phonetically [gate] < Zulu *kade* 'long ago') is in the process of being grammaticalized for pluperfect and habitual past.

Other verb inflections, which are all taken from Zulu/Nguni, are the following:

- isa* (causative), e.g., *theng-a* 'buy' versus *theng-isa* 'cause to buy, sell'
- wa* (present passive), e.g., *phek-a* 'cook'; *phek-wa* 'is cooked'
- we* (past passive), e.g., *phek-a* 'cook'; *phek-iwe* 'was cooked.'

-ela (benefactive), e.g., *theng-a* 'buy'; *theng-ela* 'buy for'-'Buy (a shirt) for (me)'

Linguistically, Fanakalo is typical of pidgins in that it cannot be classified in terms of existing language groupings; it is not quite Germanic or Nguni in structure. Its lexis and inflectional morphology stem largely from Nguni. Its syntax, however, seems to lean in the direction of the Germanic (more specifically English, rather than Afrikaans). Fanakalo is SVO in structure in main and subordinate clauses. It has none of the word-order rules of Afrikaans that place verbs at the end of subordinate clauses and at the end of main clauses that have an auxiliary in V2 position. Nor does it have the subject inversion rule of Afrikaans and of slightly archaic English that places a verb after an adverbial of time but before a subject (again in V2 position). Furthermore, there is no trace of a Zulu word-order rule that permits object pronouns to precede verbs as a clitic in unmarked (unemphatic) sentences. However, Fanakalo is not rigidly SVO insofar as it permits topic-comment order as well.

Phonetically, Fanakalo is subject to wide variation depending on the L1 of the speaker. The common core tends to use a five-vowel system (like Zulu) with two diphthongs, [ai] and [au], and to replace the clicks by velar /k/.

Bibliography

- Adendorff R (1995). 'Fanakalo in South Africa.' In Mesthrie R (ed.) *Language and social history: studies in South African sociolinguistics*. Cape Town: David Philip. 176–192.
- Cole D T (1953). 'Fanagalo and the Bantu languages in South Africa.' *African Studies* 12, 1–9.
- Mayne R (1947). *Conversational Zulu for the home*. Pietermaritzburg: Shuter and Shooter.
- Mesthrie R (1989). 'The origins of Fanagalo.' *Journal of Pidgin and Creole Languages* 4(2), 211–240.
- Mesthrie R (1998). 'Words across worlds: aspects of language contact and language learning in the eastern Cape, 1800–1850.' *African Studies* 57(1), 5–26.

Fijian

P Geraghty, University of the South Pacific, Suva, Fiji

© 2006 Elsevier Ltd. All rights reserved.

Fiji is situated in the Southwest Pacific, between the Solomon Islands and Vanuatu in the west and Tonga and Samoa in the east, but closer to the latter. There

are approximately 100 inhabited islands, with a population of 800 000.

Two hundred years ago, the vast majority of the inhabitants spoke Fijian. Today, Fijian is spoken as a first language by the indigenous population of 400 000 and by maybe 10 000 people of mixed ancestry and descendants of Solomon Islanders and other

immigrants. Fijian is also spoken in migrant communities in New Zealand, Australia, United States, and Canada.

Fijian is a continuum of some 300 distinct but related 'communalects' divided into two major subgroups, the Western in western Vitilevu (the main island) and adjoining islands, and the Eastern elsewhere. Communalects from different subgroups are not mutually intelligible, and even within each subgroup, geographical distance presents difficulties. However, the term 'Fijian' usually refers to Standard Fijian (popularly called 'Bauan'). Fijian belongs to the Central Pacific subgroup, which also includes Rotuman and all Polynesian languages. Central Pacific is a subgroup under Eastern Oceanic and then Oceanic in the Austronesian language family.

Fijian was not traditionally written and was first recorded by European visitors in the early nineteenth century. A Roman-based alphabet was devised by Methodist missionaries around 1840 and has remained in use relatively unchanged.

Although English is the main language of education, government, and commerce, Standard Fijian has a literary tradition going back over 150 years and is used to some extent in education and in the media (three radio stations and two weekly newspapers, but minimal television programming). Although Fijian speakers are mostly literate, they use their literacy very little, with literacy in English being emphasized in schools. The 1997 constitution declared Fijian one of the official languages, along with English and Hindi.

The phoneme inventory consists of 20 consonants (b [ʰb], c [ð], d [ʰd], dr [ʰr], f, g [ŋ], j [tʃ], k, l, m, n, p, q [ʰg], r, s, t, v [β], w, y, z [ʰdʒ]) and 10 vowels (a, e, i, o, u, ā, ē, ī, ō, ū). There are no consonant clusters, and

syllables are open (except in some recent loans). In writing, vowel length is not usually marked, but most modern reference works use a macron.

Grammatical functions are typically performed by affixes or pre- and postposed particles. Pronouns distinguish four persons (including first-person inclusive and exclusive) and four numbers (singular, dual, paucal, plural). Some nouns (mostly denoting body parts and kin) are suffix possessed; others suffix a possessive pronoun to a preposed marker indicating whether the possessor eats, drinks, owns, does, or is affected by the head noun. The attribute follows the noun.

There is obligatory pronominal SVO marking of subject and object within the verb phrase. It is unusual for both subject and object noun phrases to occur outside the verb phrase, but when this happens, SVO, VSO, and VOS are equally common:

era boro-ya na cauravou na no-dra vale
they paint-it the young-men the poss-their house
'The young men are painting their house'

The popular myth among linguists that 'Fijian is a VOS language' seems to have its origin in an editorial decision made by the translator of the Bible (who was not a native speaker).

Bibliography

- Capell A (1973). *A new Fijian dictionary* (4th edn.). Suva: Government Printer.
Geraghty P (1983). *The history of the Fijian languages*. Honolulu: University of Hawaii Press.
Geraghty P (1994). *Fijian phrasebook*. Hawthorn: Lonely Planet Publications.
Schütz A J (1985). *The Fijian Language*. Honolulu: University of Hawaii Press.

Finnish

F Karlsson, University of Helsinki, Helsinki, Finland

© 2006 Elsevier Ltd. All rights reserved.

Background

Finnish belongs to the (Baltic-)Finnic subbranch of the Finno-Ugric languages. The closest relatives are Karelian and Estonian. The Finno-Ugric and Samoyed languages form the Uralic language family.

In 2003, the number of Finnish speakers in Finland was 4.8 million, 92% of the population. Abroad, more than 1 million people speak Finnish (or are

descendants of Finnish immigrants), especially in Sweden (300 000), the United States (600 000), Canada, and Australia. Finnish is one of the two national languages of Finland (the other is Swedish). Finnish obtained its position as national language in 1863 and ultimately 1902. Finnish has been used in writing since the appearance of the first parts of the Bible translation in the 1540s.

Phonology

Finnish has 8 vowel and 13 consonant phonemes, /i e æ y ø u o a/ and /p t k d s h v j l r m n ŋ/. /b g/ occur

only in recent borrowings. /d/ is marginal because it occurs only as a product of morphophonological processes (consonant gradation).

Finnish stress is fixed on the first syllable. The quantity distinction is effectively phonemic. Both vowels and consonants can be phonemically short and long, and they combine with one another with few restrictions in both stressed and unstressed syllables, for example, *tuli* 'fire', *tuuli* 'wind', *tulli* 'customs', *tule* 'come!', *tulee* 'comes', *tuulee* '(the wind) blows'.

There are 16 diphthongs such as /ai æi ei oi ui ou æy ey ie uo yø/. The canonical structure of words is bisyllabic; the monosyllables can be counted in the tens.

Vowel harmony is a constraint on stems and suffixes. The vowels form three groups, the harmony vowels /y ö ä/ (front) and /u o a/ (back) plus the neutral vowels /i e/. The three vowel pairs from the harmony sets are often denoted by morphophonemic symbols /U, O, A/. Vowels from the front and back harmony sets cannot co-occur in native words whose vowels are drawn either from /i e ä y ö/ or /i e u o a/. Suffixes with harmony vowels have one front and one back variant occurring after front and back stems, respectively. Stems with neutral vowels count only as front. Thus (INE = inessive case):

talo-ssa
house-SING.INE
'in (a/the) house'

kylä-ssä
village-SING.INE
'in village'

venee-ssä
boat-SING.INE
'in boat'

Finnish orthography is often commended for being among the most efficient in the world, in the sense that it is almost perfectly phonemic. Each phoneme has its own unique letter, with the sole exception of /ŋ/ for which /ŋŋ/ is written <ng>. The phonemic perfection of Finnish orthography is true with respect to the careful normative pronunciation of the standard language. However, present-day colloquial Finnish has strayed from this ideal due to many contractions and elisions.

Morphology

Finnish is a suffixing language with an elaborate morphology. Nominals (nouns, adjectives, pronouns, and numerals) are inflected for number, case, and possessive. There are two numbers, fourteen cases, and five possessive morphemes, occurring as classes

in this morphotactic order. Here are some examples of inflected Finnish nouns.

talo
house-SING.NOM
'house'

talo-t
house-PL.NOM
'houses'

talo-ssa
house-SING.INE
'in (a/the) house'

talo-i-sta-ni
house-PL-EL-POSS.1.SING
'out of my houses'

talo-o-nne
house-SING.ILL-POSS.2.PL
'into your house'

Finite verb forms are inflected for indefinite (called passive in traditional Finnish grammar), tense and mood (belonging to the same morphotactic position because tenses and moods are mutually exclusive), and person. There are two simple tenses, present and past, and two composite ones, perfect and pluperfect. There are four moods: indicative, conditional, potential, and imperative. There are three grammatical persons in the singular and the plural, plus a fourth-person linking up with the indefinite.

sano-n
say-PRES.INDIC.1.SING
'I say'

sano-i-n
say-PAST-1.SING
'I said'

sano-isi-mme
say-COND.1.SING
'we would say'

sano
say-IMP.2.SING
'say!'

sano-kaa-mme
say-IMP-1.PL
'let us say!'

sano-ta-an
say-INDEF.PRES-4
'one says, people say'

Nonfinite verb forms (i.e., infinitives and participles) are inflected for indefinite, nonfiniteness, number, case, and possessive (INE = inessive case):

sano-a
say-INE.NOM
'to say' (infinitive I in traditional Finnish grammar)

sano-e-ssa-nne
 say-INF-INE-POSS.2.PL
 'when you are saying' (infinitive II)

sano-v-i-ssa
 say-PRES.PART-PL-INE
 'in the saying (ones)' (present participle)

sano-tta-e-ssa
 say-INDEF-INF-INE
 'when one says'

Almost every word form in Finnish, inflected or not, can be cliticized with an element from a set of five clitics with pragmatic functions. The most important one is the question morpheme *-kO/*. For example:

talo-ssa-si-ko
 house-SING.INE-POSS.2.SING-Q
 'in your house?'

Finnish lexicography as manifested in *Nykysuomen sanakirja* (Dictionary of modern Finnish, 1951–1961) postulates 82 inflectional classes for nominals and 45 for verbs; at the other extreme, a generative description might operate with none but with a wealth of ordered (morpho)phonological rules. A surface-oriented morphological approach would recognize at least 10 nominal inflectional classes and six verbal ones.

Finnish word structure is characterized by considerable allomorphy both in stems and suffixes and therefore Finnish is not a typical agglutinative language. There are tens of more or less morphologically conditioned alternations. The most profound one is consonant gradation, which concerns both nominals and verbs. The long voiceless stops /pp, tt, kk/ are shortened to [p, t, k], and the short

voiceless stops /p, t, k/ are weakened in various ways: /p/ → [m] (after /m/), /p/ → [v] (between vowels), /t/ → [d] (between vowels), /t/ → [l, r, n] (after an identical consonant), /k/ → [ŋ] (after /ŋ/), and /k/ → Ø (between vowels). These alternations are triggered by suffixation processes.

Syntax

Case marking has an important role in Finnish syntax in marking the arguments of the verb (nominative, genitive, partitive, accusative for grammatical subjects, objects, and predicate complements; and an assortment of local cases for adverbials). Due to extensive case marking, Finnish word order is free and used especially to indicate information structure, for example, subject-last for introducing new referents and leftward topicalization for linking to previous context. There are many highly productive nonfinite constructions. Premodifiers in NPs agree with the head in number and gender; the finite verb agrees with the person and number of the grammatical subject.

Bibliography

- Haarmann H (2003). 'Finnish.' In Roelcke T (ed.) *Variation Typology*. Berlin and New York: Walter de Gruyter. 866–904.
- Hakulinen A, Vilkkuna M, Korhonen R, Koivisto V, Heinonen T & Alho I (2004). *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Karlsson F (2003). *Finnish: an essential grammar*. London and New York: Routledge.
- Vilkkuna M (1989). *Free word order in Finnish: its syntax and discourse functions*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Finnish as an Agglutinating Language

F Karlsson, University of Helsinki, Helsinki, Finland

© 2006 Elsevier Ltd. All rights reserved.

A theoretically ideal agglutinative language would satisfy five criteria: there are no inflectional classes and all words of the same part of speech are inflected in the same way; there are several morphotactic positions for affixes of composite word forms, especially those of nouns and verbs; every morphological element (stem or affix) is clearly segmentable; the affixes convey one rather than several grammatical meanings;

there are no morphophonological alternations in any element due to morphological processes such as affixation. As a corollary, every element has exactly one phonological shape (disregarding low-level phonetic processes) and no fusion of several meaning elements into one unsegmentable whole.

Morphotactic Structure

Finnish is a suffixing language with a 14-member case system. As for basic inflectional and cliticized morphotactic positions, the surface structures of Finnish

nominal (nouns, adjectives, pronouns, and numerals), finite, and nonfinite verb forms are as follows (INE = inessive case, INS = instructive case, COND = conditional mood) (see Table 1, Table 2, and Table 3).

Table 1 Nominals

| Stem | Number | Case | Possessive | Clitic |
|--------------------------------------|---------|------------|-----------------|-------------|
| talo house 'in (a/the) house' | | ssa INE | | |
| talo house 'also in my house' | | ssa INE | ni POSS.1.SG | kin also |
| talo house 'from your houses?' | i PL | sta EL | si POSS.2.SG | ko Q |

Table 2 Finite verb forms

| Stem | Indefinite | Tense/mood | Person | Clitic |
|--|--------------|-------------|----------------|------------|
| sano say 'I say' | | | n PERS.1.SG | |
| sano say 'did you say?' | | i PAST | t PERS.2.SG | ko Q |
| sano say 'even if one would say' | tta INDEF | isi COND | in PERS.4 | pa even |

Table 3 Nonfinite verb forms

| Stem | Indef | Nonfinite | Number | Case | Possessive | Clitic |
|--|--------------|------------------|---------|----------------|-----------------|-------------|
| sano say 'to say' | | a INF | | | | |
| sano say 'in order for me to say' | | a INF | | kse TRANSLV | ni POSS.1.SG | |
| sano say 'when you are saying?' | | e INF | | ssa INE | si POSS.2.SG | ko Q |
| sano say 'also when one is saying' | tta INDEF | e INF | | ssa INE | | kin also |
| sano say 'saying' | | va PART.PRES | | | | |
| sano say 'said (pl.)' | | nee PART.PAST | t PL | | | |
| sano say 'of the saying (ones)' | | v PART.PRES | i PL | en GEN | | |

Inflectional Classes of Nominals

There is no consensus on how many inflectional classes there are for nominals and verbs. Traditional Finnish lexicography as manifested in *Nykysuomen sanakirja* (Dictionary of modern Finnish, 1951–1961) postulates 82 inflectional classes for nominals, whereas, at the other extreme, a generative description such as Wiik (1967) operates with none but a wealth of ordered (morpho)phonological rules. A surface-oriented morphological approach would recognize at least 10 nominal inflectional classes. The most important ones are those ending in *Bi*, *Be*, and *Bs* in the nominative singular. Four case forms (nominative, genitive, and partitive singular, partitive plural) unfold the allomorphic variation in the stems of each class, closely linked to the selection of particular ending allomorphs (NZ = nominalizer) (Table 4).

Class 1 is the largest and most productive one with the least amount of stem allomorphy, minimally only the stem vowel *Bi* alternates with *Be* before the plural *Bi*. At least 10 000 nominals are inflected according to class 1 and this is the pattern of most borrowings and other neologisms.

Classes 2 and 3 are closed: class 2 has some 220 words and class 3 around 40. Class 2 is more complex than class 1 as the stem vowel alternates also in the singular, and on top of that class 3 is more complex than class 2 by further eliding the stem vowel in SG.PART and also having alternations in the medial stem consonant. Many of the words in classes 2 and 3

are high-frequency words, for class 3, for example, *kuusi* '6,' *uusi* 'new,' *vesi* 'water,' *viisi* '5.'

For nominals in Be, class 5 is in principle closed (disregarding certain derivatives) even if it has more than 1000 members. Class 4 is small in contemporary Finnish, with less than 100 items, but this class is productive. Comparison of classes 1–3 and 4–5 discloses, not surprisingly, that the productive inflectional classes are those that have a minimal amount of stem allomorphy.

For nominals in Bs, class 6 (>4000 items) is simpler and more productive than class 7 (some 800 items). Class 8 covers a common type of derivatives.

Inflectional Classes of Verbs

Nyky-suomen sanakirja postulates 45 inflectional classes for verbs. A more generalizing approach

would do with six classes, here presented by way of four central inflectional forms, the endingless (NOM) infinitive, the first person singular present tense indicative form, the third person singular past tense form, and the past tense participle in the nominative singular form (Table 5).

Class 1 is by far the largest class with some 10 000 members. Around 2000 verbs belong to class 2, but this is presently the most productive verb class, obviously because overall there is less stem allomorphy in class 2 than in class 1; cf. especially the indefinite forms (called passives in traditional Finnish grammar) where the last morpheme is a personal ending for the indefinite 'fourth' person.

Class 4 is also a strong class with 4000 verbs. Class 3 has fewer than 20 monosyllabic verbs but around 1000 polysyllabic ones.

Morphophonological Alternations in Stems

As demonstrated by the example words of the nominal and verbal inflectional classes, Finnish word structure is characterized by considerable allomorphy, both in stems and suffixes, which detracts from the theoretical agglutinative ideal. Part of the allomorphy is most conveniently described in terms of item-and-arrangement morphophonological alternations, partly in terms of item-and-process directional rules.

Vowel harmony is an overriding constraint on stems and suffixes. The Finnish vowels form three groups, the harmony vowels /y ö ä/ (front) and /u o a/ (back) plus the neutral vowels /i e/. The three vowel pairs from the harmony sets are often denoted by morphophonemic symbols: U, O, A. Vowels from the front and back harmony sets cannot co-occur in

Table 4 Inflectional classes of nouns

| Class | SG | SG.GEN | SG.PART | PL-PART |
|-------|----------------------------|-------------|-------------|---------------|
| (1) | lasi <i>glass</i> | lasi-n | lasi-a | lase-j-a |
| (2) | ovi <i>door</i> | ove-n | ove-a | ov-i-a |
| (3) | käsi <i>hand</i> | käde-n | kät-tä | käs-i-ä |
| (4) | nalle <i>bear</i> | nalle-n | nalle-a | nalle-j-a |
| (5) | vaje <i>lack</i> | vajee-n | vaje-tta | vaje-i-ta |
| (6) | varis <i>crow</i> | varikse-n | varis-ta | variks-i-a |
| (7) | vieras <i>guest</i> | vieraa-n | vieras-ta | viera-i-ta |
| (8) | rakka-us <i>love-NZ</i> | rakka-ude-n | rakka-ut-ta | rakka-uks-i-a |

Table 5 Inflectional classes of verbs

| Class | INF.NOM | PRES.INDIC.1.SG | PAST.3.SG | PAST.PART.SG.NOM |
|-------|------------------------------|-----------------|-----------|------------------|
| (1) | sano-a <i>say</i> | sano-n | sano-i | sano-nut |
| | anta-a <i>give</i> | anna-n | anto-i | anta-nut |
| (2) | hala-ta <i>embrace</i> | halaa-n | hala-si | halan-nut |
| | kara-ta <i>escape</i> | karkaa-n | karka-si | karan-nut |
| (3) | saa-da <i>get</i> | saa-n | sa-i | saa-nut |
| | haravoi-da <i>harvest</i> | haravoi-n | haravo-i | haravoi-nut |
| (4) | nous-ta <i>rise</i> | nouse-n | nous-i | nous-sut |
| (5) | lämme-tä <i>warm up</i> | lämpene-n | lämpen-i | lämmen-nyt |

native words whose vowels are drawn either from /i e ä y ö/ or /i e u o a/. Suffixes with harmony vowels have one front and one back variant occurring after front and back stems, respectively. Stems with neutral vowels only count as front. Thus:

talo-ssa
house-SG.INE
'in (a/the) house'
kylä-ssä
village-SG.INE
'in village'
venee-ssä
boat-SG.INE
'in boat'
auto-lla
car-SG.ADESS
'by car'
pyörä-llä
bike-SG.ADESS
'by bike'
veitse-llä
knife-SG.ADESS
'with knife'
tule-vat
come-PRES.INDIC.3.PL
'(they) come'
määrää-vät
decide-PRES.INDIC.3.PL
'(they) decide'
mene-vät
go-PRES.INDIC.3.PL
'(they) go'
anne-ta-an
give-INDEF-4
hala-ta-an
embrace-INDEF-4

Vowel harmony is basically a phonological phenomenon. There are tens of other, more strongly morphologically conditioned alternations. The most profound one is consonant gradation, which concerns both nominals and verbs. Under complicated phonological and (partly opaque) morphological conditions, the long voiceless stops *pp*, *tt*, *kk* (which constitute two-phoneme combinations) are shortened to *p*, *t*, *k*, and the short voiceless stops *p*, *t*, *k* are weakened in various ways, e.g., *p* → *m* (after *m*),

p → *v* (between vowels), *t* → *d* (between vowels), *t* → *l*, *r*, *n* (after an identical consonant), *k* → *ŋ* (after *ŋ*), *k* → zero (between vowels). The weak grade is triggered by suffixation processes, in particular the occurrence of a suffix closing the syllable in the beginning of which the strong grade occurs (Table 6).

However, the weak grade also occurs in many purely morphological contexts without suffixes closing the stem syllable, e.g., in certain imperative and indefinite verb forms and in the base forms (nominative singulars) of nouns belonging to inflectional class 5: *ker.ro* tell.IMP.2.SG, *ker.ro.ta-an* tell-INDEF-4, *sa.de* rain.SG.NOM, *sa.tee-n* rain-SG.GEN.

About 20% of the Finnish vocabulary is subject to consonant gradation: 8000 nouns, 1000 adjectives, and 6000 verbs. Of the 1000 most frequent words, 30% participate in consonant gradation. This is a profound characteristic of Finnish.

Another set of typical morphophonological alternations are the vowel mutations in front of certain suffixes starting with *Bi*. Long stem vowels are shortened, stem diphthongs are simplified, stem-final short *Bi* changes to *Be*, stem-final short *Ba* might change to *Bo*, etc.:

maa
country-SG.NOM
ma-i-ssa
country-PL-INE
'in (the) countries'
tie
road-SG.NOM,
te-i-llä
road-PL-ADESS
'on (the) roads'
lasi
glass-SG.NOM
lase-i-ssa
glass-PL-INE
'in (the) glasses'
pila
joke-SG.NOM
pilo-i-ssa
joke-PL-INE
'in (the) jokes'

Many word forms simultaneously display several alternations, e.g., *virka* job.SG.NOM, *viro-i-ssa*

Table 6 Consonant gradation

| | | | |
|----------------------------------|----------------------------------|---------------------------------|-------------------------------------|
| kaup.pa shop-SG.NOM | kau.pa-n shop-SG.GEN | kau.pa-s.sa shop-SG.INE | kauppa-a shop-SG.PART |
| sil.ta bridge-SG.NOM | sil.la-n ridge-SG.GEN | sil.la-l.la ge-SG.ADESS | sil.ta-an ge-SG.ILL |
| ker.to-o tell-PRES.INDIC.3.SG | ker.ro-n tell-PRES.INDIC.1.SG | ker.ro-i-m.me tell-PAST-1.PL | ker.to.-vat tell-PRES.INDIC.3.PL |

Syllable boundaries are indicated by periods.

job-PL-INE ‘in (the) jobs’ (gradation and mutation). The nominal inflectional class 3 is particularly complex. On top of gradation and mutation (deletion of *Be*) there is assibilation of stem-internal *BtB* in SG.NOM and deletion of *Be* also in SG.PART: *käsi* hand.SG.NOM, *käte-en* hand-SG.ILL ‘into (a/the) hand’, *käde-llä* hand-SG.ADESS ‘with (a/the) hand’, *kät-tä* hand-SG.PART ‘hand’ (e.g., as direct object in negated clauses), *käs-i-ä* hand-PL-PART ‘(some indefinite) hands’.

Morphophonological Alternations in Suffixes

Suffixes might lose their final consonant in front of a consonant starting the next ending, or assimilate their first consonant to the last consonant of the preceding stem. Consonants are deleted in front of possessive suffixes: *maa-han* country-SG.ILL ‘into (a/the) country’, *maa-ha-mme* country-SG.ILL-POSS.1.PL ‘into our country’, *talo-n* house-SG.GEN, *talo-t* house-PL.NOM, *talo-mme* house-SG.NOM. POSS.1.PL ~ *house-SG.GEN.POSS.1.PL* ~ *house-PL.NOM.POSS.1.PL*. Note the three-way ambiguity arising due to consonant deletion in a form like *talo-mme* expressing nominative singular, genitive singular, and nominative plural.

The illative case has an extreme number of allomorphs. It has four basic suppletively related allomorphs, ‘quasimorphemes,’ occurring in different phonologically determined contexts: *BVn*, *BhVn*, *Bseen*, *Bsiin*. The morphophoneme *V* is realized by reduplication as a copy of the preceding vowel. In addition, the final consonant may be deleted before possessives. Therefore the illative has no fewer than 36 allomorphs:

talo-on
house-SG.ILL
 ‘into (a/the) house’
talo-o-mme
house-SG.ILL-POSS.1.PL
lasi-in
glass-SG.ILL
lasi-i-mme
glass-SG.ILL-POSS.1.PL
maa-han
country-SG.ILL
maa-ha-mme
country-SG.ILL-POSS.1.PL
puu-hun
tree-SG.ILL
puu-hu-mme
tree-SG.ILL-POSS.1.PL, etc.

Many suffixes have several allomorphs, each linked to specific stem selection criteria and/or inflectional

classes. For example, the partitive allomorph *BttA* only goes with nominal inflection class 5, *BtA* with classes 3, 6, 7, 8, and *BA* with classes 1, 2, and 4.

Fusion and Polyfunctional Suffixes

There are several polyfunctional suffixes where the grammatical functions cannot be segmented, e.g., *Bt* (nominative plural), *Bten* (genitive plural, the segmentable genitive plural has the plural marker *Bi*), *Bine* (comitative singular or plural), *-ttU* (indefinite fourth person perfect participle), *BkAA* (imperative 2.PL).

In colloquial Finnish, there is a tendency to drop *Bi* in unstressed diphthongs. When this *Bi* is the past tense marker, the mutated stem vowel becomes the only marker of the past tense function: *anta-a* give-PRES.INDIC.3.SG, *anto-i* give-PAST.3.SG, (colloquial) *anto* give-PAST.3.SG.

Conclusion

Karlsson (1983) distinguished 45 different morphophonological alternations. They create massive allomorphy in both stems and endings. A two-way dependency exists between many stems and endings: certain stems take certain endings only, and certain endings go only with certain stems. This mutual boundedness implies that Finnish word forms are highly cohesive, a property that is amplified by vowel harmony stretching over the whole (uncompounded) word, and further amplified also by the fixed initial stress.

Thus, Finnish departs from the theoretically ideal agglutinative type in some respects, as regards the occurrence of nominal and verbal inflectional classes, allomorphy among the affixes, morphophonological alternations, endings expressing composite grammatical functions, and fusion of certain grammatical elements.

Bibliography

- Hakulinen A, Vilkkuna M, Korhonen R, Koivisto V, Heinonen T & Alho I (2004). *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Karlsson F (1983). *Suomen kielen äänne- ja muotorakenne*. Juva: WSOY.
- Karlsson F (2003). *Finnish: an essential grammar*. London/ New York: Routledge.
- Paunonen H (1976). ‘Allomorfien dynamiikkaa.’ *Virittäjä* 80, 102–117.
- Wiik K (1967). *Suomen kielen morfofonemiikkaa*. Turku: Turun yliopiston fonetiikan laitoksen julkaisuja 3.

Flores Languages

M Donohue, National University of Singapore, Singapore

© 2006 Elsevier Ltd. All rights reserved.

The Austronesian languages of Flores, in southern Indonesia, all belong to the Central Malayo-Polynesian grouping, and can be divided internally into three groups, roughly western, central, and eastern. There are 11 western languages, prominent among them Manggarai and Ngad'a, which are more closely related to Bima than to the languages of the rest of Flores. The three eastern languages, Sika, Lamaholot, and Lewotobi, are related to the languages of the Solor archipelago as far as Alor. The five central Flores languages, most prominently Ende and Li'o, are the most isolating and show a chaining relationship, with the southwestern languages changing incrementally into the northeastern varieties. The languages of Flores typically employ prenasalized and imploded/preglottalized consonants, sometimes in contrast to voiced stops. Most commonly, the languages have a five vowel system with epenthetic schwas separating illicit consonant clusters. Because these epenthetic vowels do not receive stress, the epenthesis results in apparent exceptions to the penultimate stress rule. In Palu'e [l'ama] 'rice' appears to contrast in terms of stress placement with [lɛ'ma] 'tongue,' but the difference is better analysed as an underlying difference between /lama/ and /lma/.

The languages of Flores lack extensive verbal morphology to mark voice. This is part of the general isolating tendency, although the eastern languages have verbal agreement, and most languages show varying degrees of cliticization, which show degrees of development toward agreement and case marking. The eastern language Sika, for instance, shows a full set of agreement prefixes on verbs, whereas Palu'e in the center has only one proclitic, *ak* = '1sg.subj'. The recent grammaticalization of this clitic is attested in the inability of an independent pronoun to occur in the same clause as the clitic, and the optionality of the clitic: *Aku ka lama* 'I eat rice,' or *Ak=ka lama*, but not **Aku ak=ka lama*. There are no other subject clitics in Palu'e, but there are four genitive enclitics that are used in nominalized clause: *ka=gu lama* 'my eating of the rice.' The central and western languages show voice alternations. Manggarai shows an alternation in voice that is morphologically marked in word order and the choice of VP-final subject clitic. Palu'e has an active/passive alternation with only AVP and PAV word orders marking the difference, but a variety of morphosyntactic tests showing the

changed status of the A and the P. Thus, the AVP sentence *Kita ka lama wa'a* 'We ate that rice,' contrasts with PAV in *Lama wa'a kita ka* 'That rice was eaten by us.' Tests for subject, such as modification by floating quantifiers, makes this unambiguous: the clause-final quantifier *teti'ón* 'all' in *Kita ka lama wa'a teti'ón* can only modify the subject: 'We all ate that rice,' whereas in *Lama wa'a kita ka teti'ón* can only be interpreted as 'We ate all of that rice.' Other tests support this analysis of the A in a PAV construction as oblique, and the P as subject.

Symbolism and metaphor are present in both ritual and everyday speech. This is licensed by an unusually large number of homophones, partly because of constrained phonotactic possibilities. For instance, in Palu'e the fortuitous coming-together of PAN **bənuə* > *nua* 'house' and **nuSa* > *nua* 'island' is used to enforce the sense of belonging to their island home. Another factor is the extensive precategoriality of lexical roots, such as *kɬi*, which has the referential sense 'knife,' the predicative sense 'cut off,' and the modificational sense 'severed, loose.'

Bibliography

- Arka I W (2003). 'Voice systems in the Austronesian languages of Nusantara: typology, symmetry and Undergoer orientation.' *Linguistik Indonesia* 21(1), 113–139.
- Arka I W & Kosmas J (2004). 'Passive without passive morphology? Evidence from Manggarai.' In Arka I W & Ross M (eds.) *Voice in Western Austronesian languages*. Canberra: Pacific Linguistics.
- Arndt P (1931). *Grammatik der Sika-Sprache*. Ende: Arnoldus Druckerei.
- Arndt P (1933). *Lionesisch-Deutsches Wörterbuch*. Arnoldus-Druckerei, Ende, Flores.
- Arndt P (1961). *Wörterbuch der Ngadhasprache*, volume 15 of *Studia Instituti Anthropos*. Posieux, Fribourg, Suisse.
- Baird L (2002). A grammar of Kéo: an Austronesian language of East Nusantara. Ph.D. diss., Australian National University.
- Calon L F (1890/91). 'Woordenlijste van het dialect van Sikka.' *Tijdschrift voor Nederlandsche Indië* 33, 34, 501–530, 283–363.
- Calon L F (1895). 'Bidrage tot de kennis van het dialect van Sikka.' *Verhandelingen van het Bataviasche Genootschap voor Kunsten en Wetenschappen* 50, 1–79.
- Djawanai S (1983). *Ngadha text tradition: the collective mind of the Ngadha people, Flores*. Canberra: Pacific Linguistics D-55.
- Djawanai S & Grimes C E (1995). 'Ngada.' In Tryon D T (ed.) *Comparative Austronesian dictionary*. Berlin: Mouton de Gruyter. 593–599.
- Donohue M (2005). 'The Palu'e passive: from pragmatic construction to grammatical device.' In Arka I W &

- Ross M (eds.) *Aspects of Austronesian voice systems*. Canberra: Pacific Linguistics.
- Donohue M (to appear). 'Sika notes.' In Steinhauer H (ed.) *NUSA: linguistics studies in Indonesian and other languages of Indonesia*.
- Fox J J (1988). *To speak in pairs: essays on the ritual languages of eastern Indonesia*. Cambridge: Cambridge University Press.
- Keraf G (1978). Morfologi dialek Lamalera. Ph.D. diss., Universitas Indonesia. Ende, Arnoldus.
- Lewis E D & Grimes C E (1995). 'Sika.' In Tryon D T (ed.) *Comparative Austronesian dictionary*. Berlin: Mouton de Gruyter. 601–609.
- Rosen C (1986). 'Phonemes, verb classes, and personal endings in Maumere.' In Dardjowidjojo S (ed.) *NUSA: linguistics studies in Indonesian and other languages of Indonesia* 25 (8): 39–69.
- Verheijen J A J (1967). *Kamus Manggarai* (vol. 1). The Hague: Martinus Nijhoff.
- Verheijen J A J (1970). *Kamus Manggarai* (vol. 2). The Hague: Martinus Nijhoff.
- Verheijen J A J & Grimes C E (1995). 'Manggarai.' In Tryon D T (ed.) *Comparative Austronesian dictionary*. Berlin: Mouton de Gruyter. 585–592.

Formosan Languages

M Ross, Australian National University, Canberra, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Before the arrival of the Spanish, the Dutch, and the Chinese in Taiwan in the 17th century, the island was occupied by groups of people whose present-day descendants are known as the Taiwan 'aborigines' and speak Austronesian languages (see **Austronesian Languages**). Linguists use the term 'Formosan' (after an older Portuguese term for Taiwan) for these languages, to distinguish them from 'Taiwanese', the Southern Min dialect of Chinese spoken by the majority of Taiwan's inhabitants today.

A mountain chain, rising in places to almost 4000 meters, runs down the eastern half of Taiwan, and Formosan speakers live in the valleys on both sides of the cordillera and on the narrow eastern coastal strip. Before the arrival of outsiders, the plains stretching from the mountains across to the west coast were also occupied by Formosan speakers. Although almost everyone on the plains today speaks Taiwanese, Hakka, or Mandarin (the national language), several plains groups regard themselves as aboriginals despite the loss of their language. Almost all Formosan speakers also speak Mandarin. Many of those counted as aboriginals in official listings do not speak a Formosan language or are semi-speakers, and it is difficult to know who to count as speakers and how many speakers there are. For example, there are officially 3000 members of the Thao tribe, but only fifteen of these spoke the Thao language in 2003 (Blust, 2003: 1).

The Languages Today

Blust (1999) places the 14 living languages of Taiwan in nine phylogenetic groups. He also includes in his listing certain extinct languages for which reliable materials are extant (five total; marked [E]):

- Atayalic: Atayal, Seediq
- Northwest Formosan: Saisiyat, Kulon [E], Pazeh
- East Formosan: Basay-Trobiawan [E], Kavalan, Amis, Siraya [E]
- Western Plains: Taokas-Babuza [E], Papora-Hoanya [E], Thao
- Bunun
- Tsouic: Tsou, Kanakanavu, Saaroa
- Rukai
- Puyuma
- Paiwan.

Puyuma, Paiwan, Rukai, and Bunun are single-member groups. Some languages – Atayal, Seediq, Amis, Puyuma, Paiwan, Rukai and Bunun – have significant dialect variation. Amis almost certainly has the largest number of speakers, perhaps over 100 000, and is probably the only Formosan language that need not be considered endangered. At the opposite extreme, Pazeh had one speaker in 2003. Language locations are shown on the accompanying map (**Figure 1**), but in most cases these reflect a time when there were more speakers. The recognition granted to Formosan languages under Japanese rule (1895–1945) disappeared under the Kuomintang (Nationalist) government from 1949 until 1991. Since Taiwan's first democratic elections in 1991, official attitudes have come to favor other languages beside Mandarin, but this change of heart has come too late for most of the 14 Formosan languages, and their continued use by today's younger generation is very doubtful.

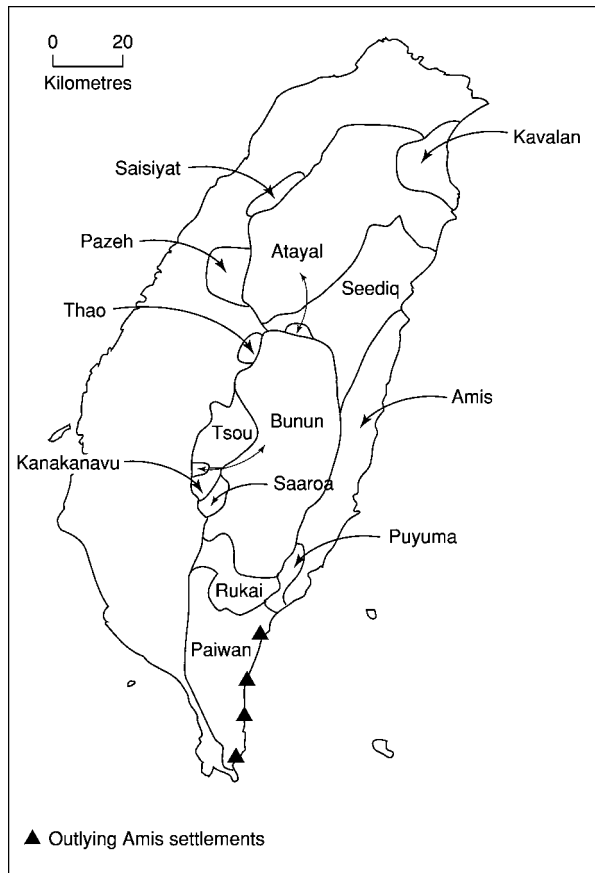


Figure 1 The Formosan languages.

History and Subgrouping

Blust's subgrouping above is based on shared phonological innovations. Each group is said to be a primary subgroup of Austronesian. There is a 10th group, Malayo-Polynesian, which includes all the other 1100 or so Austronesian languages spoken outside Taiwan (Yami of Orchid Island, politically part of Taiwan, is a Malayo-Polynesian language). If one accepts Blust's hypothesis, then the Formosan languages of Taiwan comprise 9 out of the 10 subgroups of Austronesian.

Blust's grouping is questionable in the view of some scholars because phonological innovations may diffuse across linguistic boundaries. A more conventional grouping is shown in Tsuchida's (1983) map. It has just three groups: Atayalic, Northwest Formosan (comprising Taokas-Babuza, Saisiyat, and Pazeh) and Southern Formosan, which includes three smaller groups, Tsouic, Rukai, and Paiwanic (comprising Amis, Bunun, Puyuma, and Paiwan). Other Formosan languages are ungrouped. This grouping and variations on it are open to the criticism that they

are based on shared similarities, not on shared innovations, and are therefore not subgroups in the standard sense of the linguistic comparative method.

It is possible that there will never be a full agreed subgrouping of Formosan languages. Speakers of Proto-Austronesian probably arrived in Taiwan from mainland Asia some 5000 years ago, and it is a reasonable inference that as their descendants populated Taiwan the language diversified first into a dialect network, then into individual languages. Contact between speakers of different languages would have continued, with concomitant mutual influence (and different patterns of contact at different times), leaving complex patterns of shared phonological features, lexicon, and grammatical constructions, the historical significance of which it is difficult to disentangle. Perhaps half the Formosan languages have disappeared since the early 17th century, hampering the task of historical reconstruction yet further.

There is, however, agreement among a number of scholars that Taiwan is the Austronesian 'homeland.' The reason why all Austronesian languages outside Taiwan belong to a single subgroup is that they are descended from a single language whose speakers appear, on current archaeological evidence, to have migrated southward from Taiwan to the Philippines about 4000 years ago.

The History of Formosan Language Studies

The first outsiders to study Formosan languages were probably Dutch missionaries in the early 17th century. Our knowledge of the extinct Siraya language, for example, comes from a New Testament translation and other documents printed by the Dutch (Adelaar, 2004). There is evidence that some Dutch observers recognized similarities between Formosan and Malay vocabulary, but the first explicit statement that Formosan languages belong to what we now call the Austronesian family is found in Klaproth (1824). The first modern examination of the historical position of the Formosan languages was Dyen (1963), and Dahl (1973) was the first major work on the reconstruction of Proto-Austronesian to incorporate Formosan material.

Japanese linguists and anthropologists made quite extensive studies of Formosan speakers and their languages, resulting in the publication of Ogawa and Asai (1935). The first 20th-century description of a Formosan language to appear in a Western language was Asai's (1934) account of Seediq. The first description to be published after World War II was Tung

(1964) on Tsou and two articles by Egerod (1965, 1966) that provided a sketch of Atayal grammar. This trickle has grown to a flow since the early 1970s, with grammars, dictionaries, text collections, as well as journal articles and Taiwanese M.A. theses, which mostly deal with one aspect of one language. Several single-language dialect comparisons were published by Paul Li and his colleagues at the Academia Sinica in the 1970s and 1980s. However, considerations of space allow only a restricted literature survey, and (with a few exceptions) only published book-length works in Western languages dealing with living Formosan languages are mentioned.

Languages are discussed in a roughly north-to-south order. The Atayalic languages are among the best described Formosan languages. As well as Egerod's work, there are short grammars of two Atayal dialects (Huang, 1992, 1995) as well as an important article on verbal morphemes (Huang, 2000). Egerod (1999) provided a substantial dictionary. Asai (1934), Holmer (1996), and Tsukida (2004) each sketch the grammars of Seediq dialects.

There are as yet no readily accessible descriptions of Saisiyat or Kavalan (but see Chang, 2000). However, Li and Tsuchida (2001) and Blust (2003) have provided thorough documentation of two languages on the verge of extinction, Pazeh and Thao. Ironically, for Amis, the Formosan language with the most speakers, there is only a partial description in a rather opaque framework (Chen, 1987). There is also a dictionary by Fey (1986). Bunun, another language with quite a large population of speakers, has received little attention since Jeng (1977).

Tsuchida (1976) provided sketches of the three Tsouic languages in the course of reconstructing Proto-Tsouic. Tsou had already been described by Tung (1964) and it has received continuing attention (see Zeitoun, 2004), but Kanakanavu and Saaroa are in danger of becoming extinct before being further described. Rukai was documented and described by Li (1973) and its dialects have been described in articles by Zeitoun (1997).

The grammar of one Puyuma dialect is described in Japanese with English glosses by Tsuchida (1980). Cauquelin (1991a, 1991b) provides a grammar sketch and a dictionary of another dialect. Paiwan competes with Atayal for the privilege of being the best described Formosan language, with two dictionaries (Ferrell, 1982; Egli, 2002), a grammar (Egli, 1990), and a collection of texts, some of them from Japanese work in the 1930s (Early and Whitehorn, 2003).

The documentation of Formosan languages is thus patchy, with significant gaps when it comes to

comprehensive language descriptions and dictionaries. Recently a series of short reference grammars in Chinese, of which Chang's work (2000) is an example, has been published.

The Structure of Formosan Languages

The Paiwan examples below are typical of those used to illustrate the basic structural features of Formosan clauses. Each NP is marked, either by a pronominal form or by a morpheme preceding the NP, as absolutive, genitive, or oblique (a locative marker *i* precedes or replaces the oblique in certain contexts). The verb has an affix that assigns one of four broad semantic roles to the absolutive (the 'nominative' in Formosanist parlance). This affixation is the most salient feature of Formosan languages, also shared by many Philippine languages.

q<m>alup = aken tua vavuy i (tua) gadu
<ACTOR>hunt = ABS:1SG OBL pig LOC (OBL)
mountain
'We hunt boar on the mountain.'

qalup-en nua tsautsau a vavuy i (tua) gadu tua vuluq
hunt-PATIENT GEN man ABS pig LOC (OBL)
mountain OBL spear
'The man hunts the pigs in the mountains with a spear.'

ku = qalup-an a gadu tua vavuy
GEN:1SG = hunt-LOCATION ABS mountain
OBL pig
'We hunt boar on the mountain.'

ku = si-qalup a vuluq tua vavuy
GEN:1SG = INSTRUMENT-hunt ABS spear OBL pig
'We hunt boar with a spear.'

Morphologically this is an absolutive-ergative system. The verb with ACTOR marking is intransitive, i.e., anti-passive-like, while the verb in each other example is transitive, with two core arguments: a genitive-marked actor ('genitive' because it may also serve as possessor within a NP) and an absolutive-marked NP whose role is indicated by the verbal affix.

By Formosanist convention, the verbal affix is a 'focus affix.' It would be more appropriate, however, to call it an applicative. Applicatives usually occur in nominative-accusative languages, where they change the role of the object. Here, with absolutive-ergative alignment, they change the role of the absolutive. The conventional idea that the 'focus affix' marks the semantic role of the absolutive is also suspect. On other verbs, the affixes marked as LOCATION and INSTRUMENT mark other semantic roles. The

common feature is that they progressively reduce the affectedness of the absolutive NP.

Analyzing the non-actor markers as applicatives allows us to regard the construction in which they occur as a single, 'Undergoer' voice, in opposition to the anti-passive or Actor voice. All Formosan languages except Rukai have a voice system similar to Paiwan (with some morphological differences), but its use in discourse varies from language to language (Huang, 2002). In Tsou and Puyuma, the Undergoer voice is the default in narrative discourse, and the oblique-marked Undergoer in an Actor-voice independent clause is always indefinite and often non-specific. In Paiwan and Seediq, the indefiniteness rule does not apply, and the Actor Voice behaves as an alternative default. How the speaker selects voice is not well understood.

Bibliography

- Adelaar K A (2004). 'Le siraya: interprétation d'un corpus datant du XVII^e siècle.' In Zeitoun E (ed.) *Les langues austronésiennes*. Paris: Ophrys. 123–140.
- Adelaar K A & Himmelmann N (eds.) (2004). *The Austronesian languages of Asia and Madagascar*. London: Routledge.
- Asai E (1934). 'Some observations on the Sedik language of Formosa.' *Philologia orientalis* 1, 1–84.
- Blust R A (1999). 'Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics.' In Zeitoun E & Li P J (eds.) *Selected papers from the Eighth International Conference on Austronesian linguistics*. Taipei: Academia Sinica. 31–94.
- Blust R A (2003). *Thao dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Cauquelin J (1991a). 'The Puyuma language.' *Bijdragen tot de Taal-, Land-en Volkenkunde* 147, 17–60.
- Cauquelin J (1991b). *Dictionnaire puyuma-français*. Paris: École française de l'extrême-orient.
- Chang Y (2000). *Gēmālányū cān kǎo yufǎ* [Kavalan reference grammar]. Taipei: Yuanliou.
- Chen T M (1987). *Verbal constructions and verbal classification in Nataoran-Amis*. Canberra: Pacific Linguistics.
- Dahl O C (1973). *Proto-Austronesian*. Oslo: Studentlitteratur.
- Dyen I (1963). 'The position of the Malayopolynesian languages of Formosa.' *Asian Perspectives* 7, 261–271.
- Early R & Whitehorn J (2003). *One hundred Paiwan texts*. Canberra: Pacific Linguistics.
- Egerod S (1965). 'Verb inflexion in Atayal.' *Lingua* 15, 251–282.
- Egerod S (1966). 'Word order and word classes in Atayal.' *Language* 42, 346–369.
- Egerod S (1999). *Atayal-English dictionary* (2nd edn.). Copenhagen: Royal Danish Academy of Sciences and Letters.
- Egli H (1990). *Paiwangrammatik*. Wiesbaden: Otto Harrassowitz.
- Egli H (2002). *Paiwan Wörterbuch: Paiwan-Deutsch, Deutsch-Paiwan*. Wiesbaden: Otto Harrassowitz.
- Ferrell R (1982). *Paiwan dictionary*. Canberra: Pacific Linguistics.
- Fey V (1986). *Amis dictionary*. Taipei: The Bible Society in the Republic of China.
- Holmer A J (1996). *A parametric grammar of Seediq*. Lund: Lund University Press.
- Huang L (2000). 'Verb classification in Mayrinax Atayal.' *Oceanic Linguistics* 39, 364–390.
- Huang L M (1992). *A study of Atayal syntax*. Taipei: Crane.
- Huang L M (1995). *A study of Mayrinax syntax*. Taipei: Crane.
- Huang S (2002). 'The pragmatics of focus in Tsou and Seediq.' *Language and Linguistics* 3, 665–694.
- Jeng H (1977). *Topic and focus in Bunun*. Special publication 72. Taipei: Institute of History and Philology. Academia Sinica.
- Klaproth J H (1824). 'Vocabulaire Formosan.' *Mémoires relatifs à l'Asie* 1, 354–368.
- Li P J (1973). *Rukai structure*. Taipei: Institute of History and Philology.
- Li P J & Tsuchida S (2001). *Pazih dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Ogawa N & Asai E (1935). *Gengo ni yoru Taiwan takasago-zoku densetsu-shū* [The myths and traditions of the Formosan native tribes (texts and notes)]. Taihoku [Taipei]: Institute of Linguistics. Taihoku Imperial University.
- Tsuchida S (1976). *Reconstruction of Proto-Tsouic phonology*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Tsuchida S (1980). 'Puyuma-go (Tamarakao hōgen) goi: fu gohōgaisetsu oyobi tekisuto' [Puyuma (Tamalakaw dialect) vocabulary: with grammatical notes and texts]. In Kuroshio Bunka no Kai [Black Current Cultures Committee] (eds.) *Kuroshio no minzoku, bunka, gengo* [Ethnology, cultures and languages along the Black Current]. Tokyo: Kadokawa Shoten. 183–307.
- Tsuchida S (1983). 'Austronesian languages in Taiwan (Formosa).' In Wurm S A & Hattori S (eds.) *Language atlas of the Pacific area, Part 2: Japan area, Taiwan (Formosa), Philippines, mainland and insular South-East Asia*. Canberra: Australian Academy of the Humanities and Japan Academy. Map 30.
- Tsukida N (2004). 'Seediq.' In Adelaar & Himmelmann (eds.) 291–325.
- Tung T (1964). *A descriptive study of the Tsou language*. Taipei: Institute of History and Philology.
- Zeitoun E (1997). 'Coding of grammatical relations in Mantauran.' *Bulletin of the Institute of History and Philology* 68, 249–281.
- Zeitoun E (2004). 'Tsou.' In Adelaar & Himmelmann (eds.) 259–290.

Franglais

P Rowlett, University of Salford, Salford, UK

© 2006 Elsevier Ltd. All rights reserved.

The term *franglais* encompasses combinations of French (*français*) and English (*anglais*) of various kinds. First, it refers to any variety that has developed naturally as a mixture of the two languages as a result of long-standing contact. Such varieties are spoken, for example, in New Brunswick (Canada) and northern Maine (United States). Second, the term refers to code switching between the two languages in what in some cases are again long-standing bilingual or diglossic settings. This occurs, for example, in Quebec (Canada), where, especially in Montreal since the 1960s, Anglophones frequently switch to French midsentence, just as Francophones would switch to English. Finally, *Franglais* refers to the phenomenon whereby native English or French speakers pepper their speech with lexis from the other language for humorous effect, to show off, or because of gaps in their native lexis. The use of French words in English was caricatured by Miles Kington's 'Parlez-vous Franglais?' column in the British satirical periodical *Punch* from the late 1970s and subsequently in a series of books. The reverse phenomenon, which will be our concern here, is taken much more seriously and is most closely associated with Sorbonne philosopher René Étiemble's vitriolic *Parlez-vous franglais?*, first published in 1964.

Franglais is primarily a lexical innovation: new lexis has been absorbed and existing lexis is being used differently. Francophones have enriched their lexis by (a) adopting English word forms, especially from commerce (*le shopping*, *le business*, *le discount*, *le leasing*, *le sponsor*); (b) adapting English lexis (*le self* 'self-service cafeteria', *le parking* 'parking lot', *le dressing* 'dressing room', *le loft* 'loft apartment'); and (c) innovating pseudo-English lexis (*le starter* 'choke', *le footing* 'jogging', *le baby-foot* 'table football', *le pressing* 'dry cleaner's', *le lifting* 'face-lift', *le brushing* 'blow-dry', *le planning* 'schedule', *le recordman* 'record holder', *le rugbyman* 'rugby player', *le scratch* 'Velcro').

In addition to embracing English word forms, Francophones have changed the way they use French forms, apparently under the influence of English, either as a faux ami (*un ordre* for *une commande* 'an order', *une opportunité* for *une occasion* 'an opportunity', *du matériel* for *du tissu* 'fabric') or by calquing the syntax of English expressions (*nourrir le chat* 'to feed the cat' for *donner à manger au chat*, *rejoindre l'armée* 'to join the army' for *s'inscrire à*

l'armée, *parler sur le téléphone* 'to speak on the telephone' for *parler au téléphone*). According to Duneton (1999: 121), this usage would have shocked just a few decades ago but is now banal.

However, *Franglais* is alleged to have had influence beyond lexis, too, for example, in the use of unnecessary determiners in appositive contexts (*Berlin, la capitale de l'Allemagne*); pronominal adjectives (*la composition scientifique* > *la scientifique composition*); adjectives as adverbs (*écrire économiquement* > *écrire économique* 'to write economically'). Increased use of passives has also been blamed on English influence. Finally, although *Franglais* lexis often submits to French stress patterns (breath-group final rather than lexical or foot-based) and orthography (*bouledozère* 'bulldozer', *boum* 'boom', *cédérom* 'CD-ROM'), *Franglais* has had a minor impact on the phonology and orthography of the language, with increased occurrence of the phones [ʔ] and [ɕ] (otherwise found in borrowings, only) and the graphemes *k*, *y*, and *w*. (Arguably, though, a much more apparent, economy-driven impact on French orthography is being caused by SMS messaging, with unpronounced graphemes dropped (*tabac* > *taba*, *comme* > *kom*), the representation of vowels and consonants simplified (*beau* > *bo*, *qui* > *ki*), and single graphemes used to represent whole words (*c'est/ces/ses* > *c*).

Étiemble's thesis is that, because of its scale, rather than being part of a natural process of dynamic development, the one-way influence of English on French taking place from the mid-20th century onward is, instead, doing unwelcome and irreparable damage to the core of the French language, culturally if not purely linguistically. Duneton (1999) has reached the same conclusion. Whether or not Étiemble is right depends on one's perspective. If one sees language as a mere vehicle for communication, then openness on the part of one linguistic community to the lexical resources of the language of another can only increase expressive power. Contact and mutual influence between Francophones and Anglophones have a history going back a millennium, and the languages are widely believed to be all the richer as a result. More recent longitudinal work by Shana Poplack on the effect of English–French contact on the French language spoken around Ottawa and Hull (Canada) suggests that French has been enriched by the contact, rather than impoverished. If, in contrast, like Étiemble, one sees language as having a role to play in articulating a speech community's social and cultural identity, then excessive influence of English on French arguably undermines French (or at least Francophone) identity. Distinguishing French lexical

influence on English from English lexical influence on French, for example, Duneton (1999: 115ff) reported the view that the former reflects middle-class snobbery, the latter, working-class ignorance. The desire to combat this perceived ignorance underlies much of the corpus planning pursued in France in recent decades, with official dictionaries cataloguing approved French alternatives to anglicisms and legislation guaranteeing the use of the French language in a number of contexts.

There are various reasons to believe that the stance adopted by Étiemble is more emotional than rational and similarly Duneton, one of whose grounds for concluding that the heart of the French language is under threat is the observation that Francophones now say *Oops!* – apparently unheard of in 1979 – rather than *Hou là là!* when they drop something! It is likely, for example, that Étiemble overestimated the scale of lexical Franglais. Leeman-Bouix (1994: 29–30) cites a 1965 study that counted 694 English-origin words in the 100 000 lexemes of French (0.7%). Similarly, Étiemble was wrong to claim that Franglais lexis was unwarranted due to the availability of French synonyms. For example, although *faire du shopping* may be as perfect a synonym of *faire des courses* as it is possible to get, Leeman-Bouix (1994: 137–138) justified the use of *challenger* over *adversaire* ‘opponent’ in TV game shows precisely because it includes the notion of challenge, absent from *adversaire*, and the use of *casting* instead of *distribution* because of the ambiguity of the latter term in the cinema context. Finally, Étiemble was in fact mistaken in some of the claims of English influence he made, for example, *examen-fenêtre* ‘the act of examining an object by the window’, blamed by Étiemble (1973: 195) on unwelcome English influence, despite the absence of any evidence to support such an accusation (Goosse, 2000: 131). It is thus difficult to disagree with Hagège (1987: 52) that “English has not impacted upon the hard core of the French language” and Goosse (2000: 141) that the unmarked variety, or varieties, of French is following its course.

A more plausible take on the underlying concerns of Étiemble and other purists hangs on yet another role of language, namely, its at times near-isomorphic relationship with, and reflection of, the status of a nation. There is no need to rehearse here the geopolitics of the 20th century and the way the role of

France on the international stage diminished after World War I and particularly World War II. In such a context, Étiemble’s talk of an attack by Anglo-American cultural imperialism and Duneton’s discourse of an infestation of teeming ants is hardly surprising. Franglais is a rude reminder of the loss of French prestige on the world stage. Thus, the concerns expressed by Étiemble, as well as the status-planning policies embraced by successive French governments since the mid 1960s, not to mention virulent anti-Americanism, can be seen as “a massive overreaction at a time when France’s political identity was being redefined” (Battye *et al.*, 2000: 44). Significant in this respect is a contrast between the level of negative reaction to Franglais in France and its broad acceptance in Quebec, for example, where, if anything, the influence of English is even greater.

Bibliography

- Battye A, Hintze M-A & Rowlett P (2000). *The French language today: a linguistic introduction* (2nd edn.). London: Routledge.
- Bouchard C (1999). *On n’emprunte qu’aux riches: la valeur sociolinguistique et symbolique des emprunts*. Paris: Fides.
- Duneton C (1999). *La mort du français*. Paris: Plon.
- Étiemble R (1964, 1973, 1991). *Parlez-vous franglais?* Paris: Gallimard.
- Goosse A (2000). ‘Évolution de la syntaxe.’ In Antoine G & Cerquiglini B (eds.) *Histoire de la langue française 1945–2000*. Paris: CNRS. 107–145.
- Hagège C (1987). *Le français et les siècles*. Paris: Odile Jacob.
- Hudelist R (1991). *La soupe anglaise*. Paris: Éditions de Fallois.
- Kenec’hdu T (1984). *Avatars du français: de Rivarol aux néo-linguistes, 1784–1984*. Paris: Tequi.
- Lamy M-N (1996). ‘Franglais.’ In Graddol D, Leith D & Swann J (eds.) *English: history, diversity and change*. London: Routledge.
- Leeman-Bouix D (1994). *Les fautes de français existent-elles?* Paris: Seuil.
- Lenoble-Pinson M (1994). *Anglicismes et substituts français*. Paris: Duculot.
- Thody P, Evans H & Pepratz-Evans M (1995). *Le franglais: forbidden English, forbidden American: law, politics and language in contemporary France: a study of loan words and national identity*. London: Athlone.

French

J C Smith, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

French is a member of the Romance group of Indo-European languages. It forms part of the Gallo-Romance sub-group, along with Occitan and the transitional varieties labeled Franco-Provençal. Modern French can ultimately be traced back to the Latin of northern Gaul, a Latin that was significantly modified by contact with the language of the pre-Roman Celtic inhabitants of the area and the language of the Germanic invaders who occupied the region after the fall of the Roman Empire. Its more immediate ancestor is the medieval dialect of Paris (labeled *fran-cien* by 19th-century scholars), which, as the speech of the major economic and cultural center first of the Île de France and then of a progressively larger administrative unit created by conquest and royal marriages, enjoyed particular prestige; however, in the course of becoming a national standard, this variety was additionally subjected to complex processes of leveling and koineization resulting from large-scale migration to the capital.

Number of Speakers and Geographical Distribution

French is currently spoken (as a first or near-native second language) by approximately 100 000 000 people. In Europe it is the official language of France (population approximately 60 000 000), and is also spoken in the contiguous areas of southern Belgium (roughly 5 000 000 speakers), Luxembourg (500 000), western Switzerland (1 500 000), and the Val d'Aosta region of Italy (35 000). It is one of the two national languages of Canada, where it is the native tongue of almost 7 000 000 people, or between one-fifth and one-quarter of the population – at the provincial level, it is the official language of Québec and has co-official status with English in New Brunswick; smaller numbers of speakers are found elsewhere in the maritime provinces, and in Ontario, Manitoba, and Saskatchewan. French is the official language of the French Overseas Territories (most of which are situated in the Caribbean or in the Indian or Pacific Oceans). It is also the official language of several former French colonies, especially in West, North, and Central Africa, although many inhabitants of these countries do not in fact speak French, and, for the majority of those

who do, it has the status of a second language. French dialects that differ significantly from the standard language survive in northern France (Picard) and southern Belgium (Walloon), as well as in Normandy. A Norman French dialect is also found in the Channel Islands of Guernsey and Jersey, although the numbers of speakers are fast declining. For two hundred years, from the early 18th century to the early 20th, French was the unrivaled language of international (and especially diplomatic) communication. It also played a crucial role in the establishment and subsequent functioning of the institutions from which the European Union was to emerge. In addition, it has been the vehicle of one of the world's major literatures for over a thousand years. For all of these reasons, French has assumed an importance that transcends the number of its native speakers.

Phonetics and Phonology

A maximal phonemic inventory of contemporary standard European French would include the following: 16 vowels (a front unrounded series /a ε e i/; a front rounded series /œ ø y/; a back series consisting of unrounded /ɑ/ and rounded /ɔ o u/; four nasal vowels /ā ɔ̄ ē œ̄/; and the unspecified vowel schwa /ə/); three glides /j ɥ w/; two liquids /l r/, the former a voiced lateral, the latter usually realized as a voiced uvular fricative [ʁ] or trill [R]; and 16 consonants (bilabial, dental, and velar plosives, voiced and voiceless /p b t d k g/; labio-dental, alveolar, and palatal fricatives, voiced and voiceless /f v s z ʃ ʒ/; and four nasals /m n ɲ ŋ/ – the last found only in some speakers' pronunciation of English loan-words such as *parking*). However, this inventory represents an idealization, and few, if any, speakers exemplify the full system. In particular, a number of vocalic distinctions (especially amongst mid-vowels) have, for many speakers in many contexts, been either obliterated or reduced to phonetically predictable oppositions. While present phonetically, vowel length has ceased to be phonologically significant for virtually all speakers.

The phonetics and phonology of Canadian French differ in significant ways from those of the European language, especially as regards vowels. Vowel length is more salient, and has phonemic value in some contexts (compare *mettre* [mɛtʁ] 'put' vs. *maître* [mɛ:tr 'master']; long vowels may become diphthongs. Nasal vowels are less nasal, and have chain-shifted. Short high vowels in closed syllables

become lax. Unstressed high vowels in medial syllables may be devoiced or disappear completely.

A strong (although not absolute) phonotactic principle precludes sequences of more than two consecutive consonants (or, utterance-initially, more than a single consonant), excluding syllable-initial [s] and liquids following an obstruent at the end of a cluster. This principle correlates with the presence or absence of schwa in many contexts; in informal speech, schwas will be deleted unless or until such deletion would give rise to an unacceptable sequence of consonants.

The most important and characteristic French sandhi phenomenon is *liaison*, in which a consonant that is not usually present on the surface in other contexts intervenes between two vowels at a word boundary and blocks hiatus. Historically, *liaison* arises from the differential disappearance of word-final consonants, which were retained longer before a following vowel than in other contexts. However, a synchronic treatment of the phenomenon must take account of frequent hypercorrections, which suggest that the *liaison* consonant is no longer underlying, but rather inserted epenthetically. In addition, *liaison* takes place with only a subset of consonants, and many (though not all) *liaisons* are highly variable and socially marked. In another sandhi phenomenon, misleadingly termed ‘aspirate h’ (it does not involve aspiration and may not even involve <h>), expected hiatus-blocking processes, such as elision and *liaison*, are themselves blocked (compare *l’héritier* [ləʁitje] ‘the heir’ vs. *le hérisson* [ləʁisɔ̃] ‘the hedgehog’, but also *le onze* [ləɔ̃z] ‘the number eleven’).

A strong preference for open syllables means that word boundaries do not always coincide with syllable boundaries. Prosodically, French is characterized by syllable-timing and by an absence of word stress, with tonic stress falling rather on the final syllable of the phonological phrase. Non-final syllables may occasionally bear stress for affective emphasis; otherwise secondary stress is assigned to them according to eurythmic principles. There is a tendency in Canadian French for some phrase-penultimate long syllables to receive stress.

Orthography

The French spelling system is far from phonographic. The 26 letters of the Roman alphabet are insufficient in themselves to represent all the phonemes of the language, and so the inventory of symbols has been augmented by digraphs and diacritics (the acute <’> and grave <`> accents, which originally distinguished

closed and open e, respectively, from schwa, and are occasionally used to distinguish between what would otherwise be homographs, e.g., *la* ‘the (fem.)’, *là* ‘there’, *ou* ‘or’, *où* ‘where’; the circumflex <^>, originally placed over a vowel to indicate the omission of a letter that was no longer pronounced, with concomitant lengthening of the vowel; the cedilla <ç>, indicating that <c> is pronounced as [s] before <a>, <o>, or <u>; the tréma <¨>, which, when placed over the second vowel in a sequence, prevents the sequence from being interpreted as a digraph). However, their usage, even in institutionalized normative orthography, is inconsistent. The influence of a late-medieval Latinizing tendency (which, moreover, was often ignorant of the true Latin etyma of French words) is still felt in forms such as *temps* [tā] (< Latin TEMPVS, compare Old French <tens>) ‘time’ and *poids* [pwa] (< Late Latin PENSVM, mistakenly traced to Classical Latin PONDVS) ‘weight’.

Morphology and Syntax

French is a fusional language. It has three grammatical persons, singular and plural number, and two genders, masculine and feminine. (Nominal case survived into Old French, but was then lost.) The second-person plural forms are also used to encode respect toward a single addressee. Synthetic tenses of the verb (here exemplified by *faire* ‘do’) are the present (*fais*), future (*ferai*), and past. A morphological distinction of aspect exists only in the past, where a perfective simple past (*fis*) contrasts with an imperfect (*faisais*). The subjunctive, perhaps best described as a ‘non-assertive’ mood, has a synthetic present (*fasse*) and imperfect (*fisse*). Other indicative tenses, such as the present perfect (*ai fait*), pluperfect (*avais fait*) and future perfect (*aurai fait*), as well as the perfect subjunctive (*aie fait*) and pluperfect subjunctive (*eusse fait*), are realized by combining an auxiliary (usually *avoir* ‘have’, although *être* ‘be’ is found with a subset of intransitive verbs) with the past participle. The imperfect and pluperfect subjunctives have disappeared from all but the most formal written registers of the language. A ‘future in the past’ (*ferais*, *aurais fait*), used in reported speech, doubles as a conditional mood, found in the apodosis of irrealis conditional sentences. This form also conveys attenuative values, such as politeness and evidentiality, and is coming to rival the subjunctive as an exponent of non-assertive modality.

French generally exhibits more analytic exponents than other Romance languages. Unlike Italian,

Spanish, and Portuguese, it has lost productive diminutive, augmentative, and superlative morphology. In everyday spoken language, the simple past tense (*fit*) has been ousted by the present perfect (*ai fait*), and there is evidence that a similar process is affecting the future, with the synthetic form (*ferai*) yielding to an analytic 'go to do' construction (*vais faire*). Save in exceptional cases, the singular/plural distinction in nouns and adjectives has become inaudible, with the loss of final [s], (although it continues to appear in writing) and the number of a noun is effectively indicated by some other item, usually a determiner.

Both determiners and adjectives agree with the noun. Determiners precede the noun, and attributive adjectives usually follow it, although they may also precede, especially when they have an affective nuance or a metaphorical interpretation. At sentence level, the basic word order of formal written French is SVO, but the frequent dislocations and topicalizations found in speech and in the informal written language suggest that French may be moving away from a word-order determined by grammatical functions, such as subject and object, toward one that reflects discourse-prominence.

Unlike most Romance varieties, French is not a pro-drop language, and an explicit subject is required in almost all non-imperative sentences. Subject-marking often involves a pre-verbal clitic pronoun; other clitics may mark the direct object, the indirect object, and some types of adjunct. The exact status of these clitics is the topic of much debate, with some scholars seeing them as developing into verbal affixes.

In more formal styles, interrogation is conveyed by inversion of the verb with a subject-clitic. Everyday language uses the interrogative particle *est-ce que* or declarative word-order with rising intonation. Very colloquial varieties, especially in Canada, may use a postverbal interrogative particle *-ti* or *-tu*.

In earlier stages of the language, negation was indicated by the preverbal negative marker *ne*. This marker could be reinforced by a positive-polarity item (thus *personne* 'person' > 'no one', *rien* 'thing' > 'nothing'), including *pas* 'step', which came to be a default filler of this slot and was semantically bleached. Nowadays, *ne* is generally absent from everyday spoken language, leaving the original reinforcing element as the sole marker of negation.

Vocabulary

Significant external influences on the vocabulary of French have included Frankish (during the early Middle Ages), medieval Latin, Italian (especially during the 16th century), and, latterly, English (giving rise to the hotly debated phenomenon of *franglais*). Colonial contact led to the absorption of several words from North African Arabic. Internal derivational processes include affixation, conversion, and back-formation. Two phonological processes that have proved lexically significant are clipping (e.g., *prof* < *professeur* 'teacher') and *verlan* (inversion of the order of syllables – the name, from *l'envers* 'the reverse', exemplifies the process). The latter phenomenon, which originated in low-socioeconomic-status immigrant areas in the northern suburbs of Paris, but which spread beyond these confines to become widespread in the speech of young people, and which has even penetrated the standard language, is a significant sociolinguistic development, while the outcome of some inversions has contributed to our understanding of French syllable structure.

Bibliography

- Ayres-Bennett W & Carruthers J (2001). *Studies in the modern French language: problems and perspectives*. London: Longman.
- Battye A, Hintze M-A & Rowlett P (2000). *The French language today* (2nd edn.). London: Routledge.
- Jones M A (1996). *Foundations of French syntax*. Cambridge: Cambridge University Press.
- Harris M (1978). *The evolution of French syntax: a comparative approach*. London: Longman.
- Lodge R A (1993). *French: from dialect to standard*. London: Routledge.
- Pope M K (1934). *From Latin to modern French with especial consideration of Anglo-Norman: phonology and morphology*. Manchester: Manchester University Press.
- Sanders C (ed.) (1993). *French today: language in its social context*. Cambridge: Cambridge University Press.
- Tranel B (1987). *The sounds of French: an introduction*. Cambridge: Cambridge University Press.
- Wise H (1997). *The vocabulary of modern French: origins, structure and function*. London: Routledge.

Fulfulde

J O Breedveld, Leiden University, Leiden,
The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The Language Name

Most Fulbe call their language Fulfulde. Other names currently used for the language are *Ful*, *Fula*, *Fulani*, *Peul*, and *Pulaar*. Fulfulde is in fact one large dialect continuum in Africa stretching over thousands of kilometers from Mauritania, Senegal, and Guinea in the west to as far as Sudan and Ethiopia in the east, and to Cameroon, the Central African Republic, and Congo in the south. The Fulbe call their own language *Pulaar* or *Pular* in the dialect areas of Fuuta Tooro (Senegal, Mauritania) and Fuuta Jaloo (Guinea, Sierra Leone) and *Fulfulde* in all the other dialect areas, such as Maasina (Mali), Liptaako (Burkina Faso), Gombe (Nigeria), and Aadamaawa (Cameroon). In English literature, *Fulani* is a name frequently used for both the people and their language; it is an English loanword from Hausa (spoken in Nigeria). In French literature, the people and the language are called *Peul*, which is a French loanword from Wolof (spoken in Senegal). In American literature, the name *Fula* is often used and is the name chosen for the language of the Fulbe by D.W. Arnott, who wrote an important reference book on this language. He argued that the term Fula “seems more appropriate as well as more euphonious than the plain stem *Ful*” (Arnott, 1970:1–2). The name *Ful* is often used in German; this name is based on the root, which is common to the name of the people (sing. *Pullo*, pl. *Fulbe*) and the different autochthonous language names (*pular*, *pulaar*, *fulaare*, *fulfulde*).

The Language and Its Speakers

Fulfulde is an Atlantic language, its closest relatives are Wolof and Sereer (Serer). Atlantic is a subbranch of the Niger-Congo language family.

The number of Fulfulde speakers is unknown, because for good reasons population counts usually do not determine the different ethnic backgrounds of people in various countries. In addition, numerous people in several areas of West Africa speak Fula as a second language. In the *Ethnologue* (Grimes, 2003), the number of speakers is estimated to be between 13 and 17 million.

The Fulbe are known especially for two features: in West Africa, they are almost the only group that has specialized in cattle herding (cf. Blench, 1999), and they have played an important role in the spread

of Islam in this part of the world (cf. Last, 1987). The importance of cattle for the Fulbe is reflected in the existence of a noun class NGE, which classifies all cow names and, depending on dialect, some other terms related to cows: These nouns are all pronominalized by the same pronoun *nge* (Breedveld, 1995a, 1995b). The involvement of the Fulbe with Islam is reflected in the language by numerous religious and other loanwords from Arabic (Labatut, 1984). In most areas, the class of learned Qur’anic teachers have their own sociolect that has incorporated a number of Arabic sounds (e.g., the velar fricative [x]).

Several Fulfulde key cultural words, such as *pulaaku* and *semteende*, have received a lot of attention in studies on Fulbe culture (cf. Stenning, 1959; Dupire, 1970; Riesman, 1977; Breedveld and De Bruijn, 1996). Often, these words are associated with a code of behavior. However, prescribed conduct in the Fulbe societies differs according to social class, and most descriptions take the highest social class – the *rimbe* or so-called noblemen – as the standard for the whole society, thus over-generalizing rules of behavior in certain contexts to the whole Fulbe society. In Mali, certain terms describing behavior seen as typical for the Fulbe are loanwords from neighboring languages (e.g., *yaage* ‘respect, restraint, avoidance behavior’ is a loanword from Soninke). This borrowing indicates that some component of these ideas is defined regionally and not ethnically.

Orthography

The Fulfulde language is written in both Arabic and Latin script. In both scripts, special conventions exist for writing lengths of vowels and consonants and for writing the prenasalized and laryngealized consonants (Ladefoged, 1964; Breedveld, 1995a). There is disagreement on the phonetic nature of the latter; some claim that these consonants are implosive (Sylla, 1982; Lex, 1987) or pre-glottalized (Klingenheben, 1963; Swift *et al.*, 1965). In 1966, experts attending a UNESCO meeting for the ‘Unification of Alphabets of the National Languages’ recommended a unified orthography for the Fulfulde language as shown in Tables 1 and 2 (Arnott, 1970).

Table 1 Orthography of Fulfulde vowels (short/long)

| | Front | Back |
|------|-------|------|
| High | i/ii | u/uu |
| Mid | e/ee | o/oo |
| Low | | a/aa |

Recommended by UNESCO, 1966.

Table 2 Orthography of Fulfulde consonants (short/long)

| | Labial | Alveolar | Palatal | Velar | Glottal |
|---------------|--------|----------|---------|--------|---------|
| Plosive | | | | | |
| Voiceless | p/pp | t/tt | c/cc | k/kk | ʔ/ʔʔ |
| Voiced | b/bb | d/dd | j | g | |
| Laryngealized | ʈ/ʈʈ | ɖ/ɖɖ | y/yy | | |
| Prenasalized | mb/mmb | nd/nnd | nj/nnj | ng/nng | |
| Nasal | m/mm | n/nn | ny/nyy | ŋ/ŋŋ | |
| Continuant | | | | | |
| Fricative | f | s | | | h |
| Glide | w/ww | | y/yy | | |
| Rolled | | r/rr | | | |
| Lateral | | l/ll | | | |

Recommended by UNESCO, 1966.

Different authors and ministries of education in different West African countries use the Fulfulde alphabet in different ways. For example in Senegal the symbol [ñ] is used instead of the digraph [ny] (e.g., Fagerberg-Diallo, 1983). Arnott (1970) has used the digraph [sh] for the phonetic symbol [ʃ] because that sound has replaced [c] in the Gombe dialect. Many authors describing the eastern Fulfulde dialect also use the letter [v] for a labiodental fricative that occurs in these dialects (Labatut, 1982; Mohamadou, 1991).

Noun Class System

All the nouns in Fulfulde are divided into groups with different grammatical markings. These groups are called 'noun classes,' and the number of noun classes varies according to dialect. For example, the Aadamaawa dialect spoken in eastern Nigeria and Cameroon has 25 classes, whereas the Maasina dialect spoken in Mali has 22. The noun classes group nouns according to semantic and formal grounds. Most nouns end in the suffix form of the noun class to which they belong. The reference marker or concord that is basically the same for all nouns of the same noun class occurs in all pronominalized forms of the noun (pronouns) and in all words that modify the noun (e.g., adjectives, demonstratives). In the following example from the Maasina dialect, Tioulenta (1986:6), writing a prize-winning novel about a young man leaving the village for the big city, used in one sentence repeatedly the concord KOY, which marks the plural diminutive. The concords reoccur with the forms *-oy*, *-woy*, and *-koy* in several pronominal and modifying words that refer to the nouns in the KOY-class.

Yogaaɓe mbittan njoppa ndew-oy
 some will leave leave behind little women-KOY
 naye-woy ngarbintoo-koy
 old-KOY begging-KOY
 Some will leave behind little old women begging

e ngor-oy maw-koy koy mbaaworaa.
 and little men-KOY old-KOY who are powerless
 and little old men who are powerless.

The allomorphy of the suffix forms (e.g., that which determines whether the form of the class suffix is *-oy*, *-woy*, or *-koy*) is still the subject of debate in Fulfulde studies (Klingenheben, 1941; Mohamadou, 1991; Paradis, 1992; Breedveld, 1995a; Gottschligg, 1997). The question of the semantic basis of the noun class system in Fulfulde has also recently received new attention (Mohamadou, 1991; Breedveld, 1995b). However, consensus has not been reached on these subjects.

Consonant Alternation

A large number of consonants in Fulfulde are subject to alternation in certain contexts. In most dialects, the verbal system shows a change in the initial consonant of the verb stem, depending on plural or singular number of the subject, and also when the pronominal pronoun follows rather than precedes the verb stem.

mi war-ii
 I come-COMPLETIVE
 I have come
 ɓe ngar-ii
 I come-COMPLETIVE
 I have come
 mande ngar-daa
 when come-you sg. COMPLETIVE
 When have you come?

In nominal stems, the initial consonant can alternate among three categories: a basic (continuant or fricative) consonant (F), a plosive consonant (P), and a pre-nasalized consonant (N). Table 3 shows the consonants that alternate in nominal stems.

Table 3 Consonant alternation in Fulfulde

| Basic (F) | w | r | y | y | w | b | d | j | g | f | s | h |
|------------------|----|----|----|----|----|----|----|----|----|---|---|---|
| Plosive (P) | b | d | j | g | g | b | d | j | g | p | t | k |
| Prenasalized (N) | mb | nd | nj | ng | ng | mb | nd | nj | ng | p | t | k |

Basically, the articulatory nature of the first consonant of a nominal depends on the noun class suffix. This distribution is shown in the following nominal paradigms derived from the verb stem *yim-* ‘to sing, recite’:

- *yim-re* poem (NDE class suffix: F initial consonant)
- *jim-e* poems (’DE class suffix: P initial consonant)
- *jim-ol* (long) song (NGOL class suffix: P initial consonant)
- *jim-i* (long) songs (’DI class suffix: P initial consonant)
- *jim-el* small song, poem (NGEL class suffix: P initial consonant)
- *njim-oy* small songs, poems (KOY class suffix: N initial consonant)
- *jim-al* big song, poem (NGAL class suffix: P initial consonant)
- *jim-eele* big songs, poems (DE class suffix: P initial consonant).

That the form of the initial consonant of nouns is determined by the class is a possible remnant of the fact that the classes that are now marked by suffixes were once prefixes (De Wolf, 1985).

There are dialect variations in the system of consonant alternations. For example, there is no initial consonant alternation in the verbal system in the dialects of Fuuta Jaloo. Additional consonants alternate in the nominal systems; for example, the dialect of Fuuta Tooro (Senegal) has an additional set ’-g-ng, and in Aadamaawa (Nigeria, Cameroon), the additional consonant alternation set v-b-mb occurs.

Focus as a Salient Feature of Fulfulde Syntax and Verbal Morphology

Most literature devoted to Fulfulde syntax also deals with its complicated verbal morphology, because both the form of the verbal conjugation and word order are determined by focus (Labatut, 1982; McIntosh, 1984). For example, in a sentence without any of the constituents in focus, the marker of the incomplete verbal conjugation is *-an*. As soon as one of the constituents is in focus, the focused constituent is placed in the first position of the sentence and the so-called relative form of the incomplete *-ata* is then used to conjugate the verb.

Baaba sood-an kaddule.
father buy-INCOMPLETE clothes
Father will buy clothes.

Kaddule baaba sood-ata.
clothes father buy-RELATIVE.INCOMPLETE
CLOTHES father will buy.

There are three voices – active, middle, and passive – marked in the Fulfulde verb, which have ramifications for the verbal morphology and constituent order. The following three sentences show combinations of the verb stem *yeggit-* ‘forget’ and the negative incomplete conjugation in the three voices.

Mo yeggit-ataa ndee innde.
he forget-NEGATIVE INCOMPLETE that name
He will not forget that name.

Ndee innde yeggit-ataako.
that name forget-NEGATIVE INCOMPLETE
MIDDLE VOICE
That name cannot be forgotten.

Ndee innde yeggit-ataake.
that name forget-NEGATIVE INCOMPLETE
PASSIVE VOICE
That name will not be forgotten.

Sentences in the passive and middle voice usually have one constituent less than sentences in the active voice. There are three paradigms of conjugational verb suffixes: each voice had its own set.

Linguistic Taboos

Part of what Fulbe consider to be proper behavior is not to say what should not be said. Certain words are taboo for all speakers, and certain names and terms of address are taboo in particular (kinship) relations.

The taboo on body part nouns has led to much dialect variation in Fulfulde. In Malim, the euphemism for the back is *caggal*, and *baawo* is considered rude. Conversely, *baawo* is the proper word in Cameroon. Because prepositions are grammaticalized from some body part terms, the same dialectal variation is replicated. In Mali, the preposition *nder* ‘in’ is not used, possibly because it is derived from the noun *reedu* ‘belly’, which is considered rude. In certain dialects, the noun class concord *ngu* is taboo because it is associated with the female genitals.

There are many taboos on names (Ameka and Breedveld, 2004). There is a general tendency to use clan names, rather than more personal first names. In certain specific kinship relations, names are replaced by other words (e.g., a child named after his or her grandmother is called *innere* 'little mother-thing').

Although many studies have been written on the Fulfulde language (cf. Seydou, 1977), descriptions of many dialects are lacking. The study of dialect comparison remains an important goal for further research.

Bibliography

- Ameka F K & Breedveld A (2004). 'Areal cultural scripts for social interaction in West African communities.' *Intercultural Pragmatics* 1/2.
- Arnott D W (1970). *The nominal and verbal systems of Fula*. Oxford: Clarendon.
- Blench R (1999). 'Why are there so many pastoral groups in East Africa?' In Azarya V et al. (eds.) *Pastoralists under pressure*. Leiden: Brill. 29–49.
- Breedveld A (1995a). *Form and meaning in Fulfulde*. Leiden: CNWS.
- Breedveld A (1995b). 'The semantics basis of noun class systems: the case of the KI and NGE classes in Fulfulde.' *Journal of West African Languages* 15(2), 63–74.
- Breedveld A & De Bruijn M (1996). 'L'image des Fulbe: analyse critique de la construction du concept de 'pulaaku'.' *Cahiers d'Etudes Africaines* 36(144), 791–821.
- De Wolf P P (1995). *English-Fula dictionary* (3 vols). Berlin: Dietrich Reimer.
- Dupire M (1970). *Organisation sociale des Peuls*. Paris: Plon.
- Fagerberg-Diallo S (1983). Introduction to Pulaar, the dialect of the Haalpulaar'en, Northern Senegal. Unpublished manuscript.
- Gottschligg P (1997). 'Nominale Morphophonologie des Ful von Klingenheben bis Paradis: Keine Rezeptionsgeschichte.' *Afrika und Übersee* 80, 61–187.
- Grimes B (2003). *Ethnologue* (14th edn.). <http://www.sil.org/ethnologue/>.
- Kirk-Greene A H M (1986). 'Maudu laawol pulaaku: survival and symbiosis.' In Adamu M & Kirk-Greene A H M (eds.) *Pastoralists of the West African savanna*. Manchester: Manchester University Press. 40–55.
- Klingenheben A (1941). *Die Suffixklassen des Ful*. (Beihefte zur Zeitschrift für Eingeborenen-Sprachen, 23.) Berlin: Dietrich Reimer.
- Klingenheben A (1963). *Die Sprache der Ful*. Hamburg: J. J. Augustin.
- Labatut R (1984). 'Les emprunts du peul à l'arabe.' In Nicolas G (ed.) *Langues arabe et langues africaines*. Paris: CILF. 41–70.
- Labatut R (1982). 'La phrase peule et ses transformations.' Ph. D diss., Université de Paris III.
- Ladefoged P (1964). *A phonetic study of West African languages*. Cambridge: Cambridge University Press.
- Last M (1987). 'Reform in West Africa: the Jihad movements of the nineteenth century.' In Ajayi J F A & Crowder M (eds.) *History of West Africa, Vol. II*. New York: Longman.
- Lex G (1987). *Etude phonétique des "glottalisées" du peul (Fouladou, Sénégal)*. Mémoire. Paris: D. R. E. A., INALCO.
- McIntosh M (1984). *Fulfulde syntax and verbal morphology*. Boston: Kegan Paul and Routledge.
- Mohamadou A (1991). 'Classificateurs et représentation des propriétés lexicales en peul, parlars d'Aadamaawa.' Ph.D. diss., INALCO.
- Paradis C (1992). *Lexical phonology and morphology, the nominal classes in Fula*. New York: Garland.
- Riesman P (1977). *Freedom in Fulani social life*. Chicago: University of Chicago Press.
- Seydou C (1977). 'Bibliographie générale du monde peul.' *Études Nigériennes* 43.
- Swift L B et al. (1965). *Fula basic course*. Washington, DC: Foreign Service Institute.
- Sylla Y (1982). *Grammaire moderne du pulaar*. Dakar: Nouvelles Editions Africaines.
- Tioulenta T (1986). 'Nafannde Arsike wittii Bolobaana.' *Jama* 3, 4–43.

This page intentionally left blank



Galician

J Beswick, University of Southampton,
Southampton, UK

© 2006 Elsevier Ltd. All rights reserved.

Language Classification

Galician is a Romance language derived from Vulgar Latin, belonging to the family of Ibero-Romance language varieties, and specifically, to the Galaico/Galician-Portuguese linguistic area. Termed *Galiza* by pro-Portuguese groups; in Galician, the language is *galego*.

Historical Overview

Galicia is the northwest region of Spain. One of three self-governing regions, the indigenous language of over 3 million people shares coofficial status with Castilian but linguistic origins with Portuguese, whose independence initiated diversification of Galician-Portuguese linguistic systems. By the 17th century, Galician varieties had lost prestige functions to Castilian, solely retained for oral intra-group purposes. However, by the 1980s a regional constitution and a written standard, the *Normas Ortográficas da Lingua Galega* (RAG & ILGA, 1995), were endorsed.

The long-standing diglossic situation with Castilian has led to lexical interference, although the phonological system also exhibits some structural borrowing: unlike Portuguese, *b* and *v* are homogeneous phonetically, there are no phonological nasal vowels or distinctive voiced sibilants, and the *ceta* [θ] is present.

Two major differences between Galician/Portuguese and Castilian:

- | | | |
|--------------|---------------------|---------------------------|
| (1) Lat. Ĕ > | Port./Gal. Cast. | terra tierra 'land' |
| Lat. Ō > | Port./Gal. Cast. | porta puerta 'door' |

- (2) Castilian monophthongization:

| | |
|------------|------------------|
| Port./Gal. | madeira |
| Cast. | madera 'wood' |
| Port./Gal. | pouco |
| Cast. | poco 'little' |

Phonological System

The three dialectal zones based on derivations of Latin -ANUM by García de Diego (1909)

- (3) eastern -ao
western -an
central -ano

may be maintained, but like the *seseo* pronunciation of *c*, the *gheada*, the (variable) voiceless continuant pronunciation of *g* in certain western dialects, confirms Zamora Vicente's (1953) two linguistic zone thesis; *galego iriense* (west) and *galego lucense* (east) (Fernández Rei, 1990):

- (4) Gal. amigo > ami[h]o
'friend' > ami[x]o
> ami[x^h]o

However, the *Atlas Lingüístico de Galego* (ILGA, 1990, 1995, 1999) highlight the continued presence of more diverse dialectal variation. For example, whereas Portuguese acquired *Ŷ* or retained *V[m]* of word-final VN sequences, and Castilian retains *V[n]*, Galician generally adopts a *V[ŋ]* resolution, but the following can also occur:

- (5) Lat. GERMANU > Port. irmão [ãw]
'brother' > Gal. irmao [aw]
(center, east)
> irmán [aŋ] (west)
> irmá [a] (northwest)
> Cast. hermano [an]

The *Normas* also require [ŋ] for intervocalic *nh*:

- (6) Gal. unha
'one' (fem. sing.)

Controversy surrounding the phonemic status of [ŋ] is based upon its syllable position (Veiga Arias, 1976; González González and González González, 1994; Beswick, 1999).

See **Table 1** for a more complete display of the Galician Phonological System.

Morphology

The tense range is similar to Portuguese: no compound tense forms, but a personal infinitive:

- (7) *comeran cando eu os chamei*
'they had eaten when I called them'

Table 1 Galician phonological system

| Oral Vowels | | | | | |
|--|------|--------------------------------|------------------|-------------------------------|--|
| Tonic | | | Non-final atonic | | Final atonic |
| a | [a] | <i>cama</i> 'bed' | [a] | <i>palabra</i> 'word' | [a] <i>casa</i> 'house' |
| e | [e] | <i>cera</i> 'wax' | [e] | <i>escrito</i> 'written' | [e] <i>lume</i> 'light' |
| | [ɛ] | <i>letra</i> 'letter' | | | |
| i | [i] | <i>clima</i> 'climate' | [i] | <i>último</i> 'last' | |
| o | [o] | <i>son</i> 'sound' | [o] | <i>época</i> 'era, period' | [o] <i>novo</i> 'new' |
| | [ɔ] | <i>forma</i> 'shape' | | | |
| u | [u] | <i>grupo</i> 'group' | [u] | <i>portugués</i> 'Portuguese' | |
| Diphthongs | | | | | |
| Falling | | | Rising | | |
| ai | [aj] | <i>pai</i> 'father' | ia | [ja] | <i>copia</i> 'copy' |
| au | [aw] | <i>auga</i> 'water' | ie | [ie] | <i>ciencia</i> 'science' |
| ei | [ej] | <i>maneira</i> 'manner' | io | [jo] | <i>milenio</i> 'millennium' |
| eu | [ew] | <i>meu</i> 'my' | iu | [ju] | <i>diurno</i> 'daily' |
| iu | [iw] | <i>partiu</i> 'he, she breaks' | ua | [wa] | <i>igual</i> 'same' |
| oi | [oj] | <i>biscoito</i> 'biscuit' | ue | [we] | <i>frecuencia</i> 'frequency' |
| ou | [ow] | <i>doutor</i> 'doctor' | ui | [wi] | <i>lingüista</i> 'linguist' |
| ui | [uj] | <i>puiden</i> 'I was able to' | uo | [wo] | <i>residuo</i> 'residue' |
| Consonants | | | | | |
| Orthographic symbol | | | | | |
| b intervocalically | | | [β] | | <i>beber</i> 'to drink' |
| b elsewhere | | | [b] | | |
| c + e, i | | | [θ] or [s] | | <i>cedo</i> 'early' |
| c + a, o, u | | | [k] | | <i>carta</i> 'letter' |
| ch | | | [tʃ] | | <i>chiste</i> 'joke' |
| d intervocalically | | | [ð] | | <i>dedo</i> 'finger' |
| d elsewhere | | | [d] | | |
| f | | | [f] | | <i>feo</i> 'ugly' |
| g | | | [g] or [h] | | <i>garfo</i> 'fork' |
| gu + e, i | | | [g] | | <i>guerra</i> 'war' |
| l | | | [l] | | <i>lei</i> 'law' |
| ll | | | [ʎ] or [j] | | <i>allo</i> 'garlic' |
| m | | | [m] | | <i>mesa</i> 'table' |
| n | | | [n] | | <i>nó</i> 'knot' |
| ñ | | | [ɲ] | | <i>viño</i> 'wine' |
| nh | | | [ŋ] | | <i>unha</i> 'one' |
| p | | | [p] | | <i>persoa</i> 'person' |
| q | | | [k] | | <i>quente</i> 'hot' |
| r intervocalically, word finally r elsewhere | | | [r] | | <i>ira</i> 'anger'; <i>ser</i> 'to be' |
| r elsewhere | | | [r] | | <i>rede</i> 'net'; <i>tenro</i> 'tender' |
| rr | | | [r] | | <i>carro</i> 'cart' |
| s | | | [s] | | <i>sabor</i> 'taste' |
| t | | | [t] | | <i>tema</i> 'theme' |
| v intervocalically | | | [β] | | <i>vivir</i> 'to live' |
| v elsewhere | | | [b] | | |
| x | | | [ʃ] | | <i>xente</i> 'people' |
| z | | | [θ] or [s] | | <i>zapato</i> 'shoe' |

- (8) para faceres iso, precisas axuda
'in order to do that, you'll
need help'

Verbal periphrases express aspectual, modal, or temporal meanings:

- (9) o neno está a chorar/ o neno está chorando
'the boy is crying'
- (10) teño traballado moito cando era xoven
'I used to work very hard when I was young'
- (11) o can foi atropelado polo coche
'the dog was run over by the car'

Words in *-n* and polysyllabic words in *-l* are variable in the plural according to the linguistic zone:

- (12) o can > os cans (west/standard)
'dog' > os cas (center)
> os cais (east)
- (13) o animal > os animás (sporadic)
'animal' > os animais (west/center:
castilianism)
> os animais (east/standard)

The second person dative pronoun *che* is syntactically optional as a solidarity dative:

- (14) dóecheme a perna
'my leg aches'

Object pronouns have a fixed order, as in Portuguese; contraction is common:

- (15) déullelo
deu + lles + o
'he gave it to them'

Third person accusative pronoun allomorphs are phonologically conditioned by final phonemes of preceding words:

- (16) o/a/os/as - default
comen o pan
'they eat the bread'
cómeno
'they eat it'

Here, *n* may be a positional allophone: velar word-finally before the article but morphologically connected to the pronoun, syllable-initial, and hence, alveolar.

- (17) lo/la/los/las - following *-r/-s*, leading to their loss:
vou beber + a
vou bebela
'I am going to drink it'

- (18) no/na/nos/nas - following a diphthong.
A widespread innovative trait:

farei o xantar
'I will make dinner'
fareino
'I will make it'

The definite article comprises two allomorphs:

- (19) o/a/os/as - default
- (20) lo/la/los/las - following *-r/-s*:
todos os días
todos + os
'everyday'

Article/preposition contractions are common:

- (21) ós luns
a + os
'on Mondays'
- (22) dunha amiga
de + unha
'of a friend'

Syntax

Pronoun order is similar to Portuguese. Simple declaratives and interrogatives:

- (23) A miña irmá tróuxome unha flor
'My sister brought me a flower'
- (24) Faláronlle da festa?
'Did they talk to him about the party?'

Negative, subordinate, complex interrogatives:

- (25) A muller non che deu as noticias
'The woman didn't give you the news'
- (26) Sabes canto me custou a casa
'You know how much the house
cost me'
- (27) ¿Quen lle deu o libro?
'Who gave him the book?'

Bibliography

- Alvarez Blanco R (1983). 'O artigo en galego. Morfoloxía.' *Verba* 10, 169–182.
- Beswick J E (1999). The velar nasal in Galicia: case-study of a linguistic variable in a bilingual community. Ph.D. Diss., University of Bristol.
- Beswick J E (2002). 'Galician language planning and implications for regional identity: restoration or elimination?' *National Identities* 4(3), 257–272.
- Carballo Calero R C (1981). *Problemas da lingua galega*. Lisbon: Sá da Costa.

- Fernández Rei F (1990). *Dialectoloxia da lingua galega*. Vigo: Edicións Xeraisde Galicia.
- García de Diego V (1909). 'Elementos de gramática histórica gallega.' [Burgos]. *Verba, Anejo* 23. Universidade de Santiago de Compostela, 1984.
- González González M & González González M (1994). 'Algunhas consideracións ó redor do /ɲ/ en galego.' Paper presented at IV Congreso da Asociación Internacional de Estudos Galegos, University of Oxford, Sept. 1994.
- Green J (1994). 'Language status and political aspirations: the case of northern Spain.' In Parry M M, Davies W V & Temple R A M (eds.) *The changing voice of Europe*. Cardiff: University of Wales Press. 155–172.
- ILGA (1990). *Atlas lingüístico galego 1: Morfoloxía verbal*. A Coruña: Fundación Barrié de la Maza.
- ILGA (1995). *Atlas lingüístico galego 2: Morfoloxía non-verbal*. A Coruña: Fundación Barrié de la Maza.
- ILGA (1999). *Atlas lingüístico galego 3: Fonética*. A Coruña: Fundación Barrié de la Maza.
- Monteagudo H & Santamarina A (1993). 'Galician and Castilian in contact: historical, social and linguistic aspects.' In Posner R & Green J N (eds.) *Trends in romance linguistics and philology 5: Bilingualism and linguistic conflict in romance*. The Hague: Mouton. 117–173.
- RAG & ILGA (1995). *Normas ortográficas e morfolóxicas do idioma galego* (12th edn.). Santiago de Compostela.
- Veiga Arias A (1976). *Fonología gallega*. Valencia: Bello.
- Zamora Vicente A (1953). 'De geografia dialectal: -ao, -an en gallego.' *Nueva revista de filología hispánica* 7, 73–80.

Gamilaraay

P K Austin, School of Oriental and African Studies, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Gamilaraay (Kamilaroi) is an Australian Aboriginal language that was traditionally spoken over a large area in the northwest of New South Wales, from the Great Diving Range, near Tamworth, north and west to the Darling and Barwon rivers. There was a range of dialect variation within this region, mostly marked by vocabulary differences, with all the local groups identified as *gamil* 'no' -*araay* 'having'. The sociolinguistic history of Gamilaraay is typical of many southeastern Australian languages.

The first recording of Gamilaraay is a short word list collected in 1832 by Major Thomas Mitchell (1839), and there is quite an amount of vocabulary material collected by local landowners in the late 19th century. The missionary William Ridley (Ridley, 1856, 1875) lived among the Gamilaraay in the 1850s and studied the language, collecting vocabulary and making simple primers and Bible translations. In 1903 the surveyor R. H. Mathews published grammatical notes and a short word list; however, the first professional recordings of the language date from 1930, when the anthropologist Norman Tindale took down words in phonetic notation and collected a short traditional narrative text (Austin and Tindale, 1986). By that time, local Aboriginal social and cultural transmission had been so disturbed by the impact of

European settlement (*see Australia: Language Situation*) that the two old men Tindale interviewed had difficulty recalling the story. In 1955, S. A. Wurm carried out extensive fieldwork in eastern New South Wales and at Boggabilla interviewed Peter Lang, who seems to have been the last fluent native speaker of the language. He died the following year. Wurm's materials included field notes and a 13-minute tape recording. In 1972–1974, Austin interviewed a large number of semispeakers who could recall up to 200 lexical items and fixed phrases from their youth, though none could produce sentences in the language. Using all the existing modern and 19th-century materials, together with comparative data from neighboring languages, one can obtain a fair but incomplete idea of the language and its structure.

From the 1940s onward, Gamilaraay ceased to be used as the main means of communication, although knowledge of words and expressions (such as plant, animal, and food names) continues until today. Since the 1990s, there has been intense local interest in the language, and strong support for its documentation and reintroduction. Austin (1992) is a bilingual dictionary that has been widely distributed; a hypertext version created by Austin and Nathan (1995) was the first fully hypertext bilingual dictionary on the World Wide Web.

As a result of local initiatives and with support from the New South Wales government, Gamilaraay is currently undergoing a language revival and is being taught both in adult education and primary school classes. A range of materials are now available on the language and its neighbor Yuwaalaraay, including a

reference dictionary (Giacon *et al.*, 2001), language lessons (Giacon, 1999), and a wordbook with accompanying music CD.

Language Relationships

Gamilaraay is closely related to its immediate western neighbors, Yuwaalaraay and Yuwaaliyaay. The languages share about 70% common vocabulary with Gamilaraay, and a similar grammatical system. Fortunately, Corinne Williams was able to study with the last two fluent speakers of these languages in 1975 and compiled a basic reference grammar and vocabulary list (Williams, 1980). There is also a large amount of tape-recorded material, collected in the 1970s by amateur linguist Janet Mathews (a relative of R. H. Mathews), that is being mined for other data.

These languages are quite clearly related to Wiradjuri (Wiradhuri), spoken over a large area of central New South Wales, and Ngiyampaa and Wayilwan, spoken along the Lachlan River (Donaldson, 1980; Austin *et al.*, 1980), and are members of a single linguistic subgroup (see Austin, 1997). This subgroup belongs to the widespread Pama-Nyungan family, which covers the southern two-thirds of Australia (see *Australia: Language Situation*).

Linguistic Characteristics

Phonology

The phonological system of Gamilaraay is typical of languages of eastern Australia, with contrastive stops at five points of articulation, a nasal for each stop position, a single lateral, a flap, a semiretroflex continuant, and two glides. **Table 1** gives the relevant consonants in their practical orthographic form. There are just three vowels: high front *i*, high back *u*, and low *a*, with a phonemic length contrast found in all syllables of words.

The general structure of Gamilaraay roots is CV(C)CV(C). Every word must begin with a consonant and end in a vowel, or *n*, *l*, or *y*. Word-initially, only nonapical stops and nasals and the two glides

are found. Word-medially, there are limited consonant clusters, primarily homorganic nasal plus stop, and apical nasal or lateral plus peripheral stop (*b* and *g*). Vowel clusters are not found. Words borrowed from English are generally restructured to meet these phonotactic constraints, e.g., *wajiin* ‘white woman’ (from ‘white gin’), *ganjibal* ‘policeman’ (from ‘constable’). Word stress is entirely predictable from the phonological shape of words: primary stress falls on long vowels or on the first syllable of a word that does not contain a long vowel. Secondary stress is on each even-numbered syllable to the left or right of the primary stress (except that final short syllables are not stressed). Examples are *gamilaraay* [g^ham^hil^har^hay] ‘Gamilaraay’, *bandaar* [ban^hdar^h] ‘kangaroo’, *thinawan* [tin^hwan^h] ‘emu’.

Morphology

Gamilaraay, like other languages of the Pama-Nyungan group, is entirely suffixing in its morphology. There are two major word classes: nominals and verbs, with nominals showing a rich system of case marking and verbs marking tense/aspect/mood and dependent clause categories. Nominals can be subdivided into substantives (which cover both noun and adjective concepts in a language like English), pronouns, and demonstratives. Minor word classes include adverbs, particles, and interjections.

Nominals in Gamilaraay inflect for case, with the syntactic functions of intransitive subject (S), transitive subject (A), and transitive object (P) showing a split-ergative pattern of syncretism in the case forms determined by animacy:

- For the first and second person pronouns, S and A fall together as a single (unmarked) form with P different, making nominative-accusative case marking.
- For third person pronouns and all other nominals, S and P fall together as a single (unmarked) form with A different, making ergative-absolutive case marking.

In addition to the three main cases (*nominative* for S, *ergative* for A, *accusative* for P), there are also the following case forms:

- *dative* marking alienable possession, and direction toward a place
- *locative* coding location in a place
- *ablative* coding direction from a place, and cause

The actual forms of the cases are affected by the phonological shape of the root, e.g., whether it ends in a vowel or not and what kind of vowel or consonant is root-final.

Table 1 Consonants

| | Bilabial | Lamino-Dental | Palatal | Apico-Alveolar | Dorsovelar |
|------------|----------|---------------|-----------|----------------|------------|
| Stop | <i>b</i> | <i>dh</i> | <i>j</i> | <i>d</i> | <i>g</i> |
| Nasal | <i>m</i> | <i>nh</i> | <i>ny</i> | <i>n</i> | <i>ng</i> |
| Lateral | | | | <i>l</i> | |
| Flap | | | | <i>rr</i> | |
| Continuant | | | | <i>r</i> | |
| Glide | <i>w</i> | | <i>y</i> | | |

Gamilaraay has a well-developed system of nominal word-building morphology that involves suffixation between the root and case inflection. Categories encoded in word-building morphology include number (plural), having (e.g., *bagaay-baraay* ‘Boggabri’ (literally ‘creek-having’)), lacking (e.g., *yuul-ngin* ‘hungry’ (literally ‘food-lacking’)).

Pronouns in Gamilaraay distinguish three persons and singular, dual, and plural number; in the first person nonsingular, there is no inclusive-exclusive contrast (unlike other Australian languages). **Table 2** sets out the basic pronoun forms. There are also bound pronouns for second person reference only; these are reduced forms of the free pronouns and are suffixed to sentence initial negative and interrogative particles only. Examples are the following:

- (1) *Yaama-nda ngalingu wuu-rri dhingga*
Q-2sg 1dl.dat give-fut meat
‘Will you give us meat?’
- (2) *Gariya-ndaali dhingga nhama dha-la*
not-2pl meat this eat-imper
‘Don’t you two eat this meat!’

Verbs morphologically distinguish between main verb and dependent verb inflections. Main verbs encode tense and mood categories, distinguishing future, nonfuture (covering past and present time reference), and imperative. Dependent verbs occur in hypotactically linked clauses and mark relative present tense (giving background information about the main clause event) and relative future tense (typically expressing purpose). The relative future is formally future plus dative case. There are four morphologically determined verb conjugations: conjugation 1 is primarily

transitive, conjugation 2 is primarily intransitive, and conjugations 3 and 4 are much smaller and have mixed transitivity. There are monosyllabic verb roots that occur in all conjugations. **Table 3** sets out the verb conjugation endings.

Verbs show productive word-building morphology, including affixes that indicate the temporal reference of an event within the tense frame of the inflected verb, e.g., *-ngayi-* indicates ‘event in the morning’, *-mayaa-* ‘event in the evening’, and aspectual, e.g., *-waaba-* ‘completive’, or directional meanings, e.g., *-uwi-* ‘back’. There are also transitivity and detransitivizing affixes, which shift conjugation and transitivity, e.g., *-ala-* ‘reciprocal’, *-ngiili-* ‘reflexive’. There are also limited category-changing processes with only nominalization marked by addition of a conjugation marker to the verb, e.g., *giili-y* ‘urine’ being productive.

The minor categories of adverb, particle, and interjection show no morphological variation.

Syntax

Like other Australian languages (see **Jiwarli**), Gamilaraay has relatively free word order and shows all possible orders of Subject, Object, and Verb, although there is a preference for A P V order; Williams (1980: 93) said this is found in 65% of examples. It also allows nouns and adjectives to be separated in the clause, with case agreement indicating which elements are constituents. Williams (1980: 96) gave the Yuwaalaraay sentence showing this:

- (3) *Buma-ay dhayin-du buyabuya dhayin*
hit-nonfut man-erg thin man
wamu-bidi-ju
fat-big-erg
‘The fat man hit the thin man’

When the adjective precedes the noun, no case marker needs to be attached to the adjective. Similarly, possessors (in dative case) may precede or follow the alienable possessed noun.

Gamilaraay interclausal syntax is relatively simple compared with some other Australian languages. Dependent clauses occur hypotactically, located on the margins of main clauses, and distinguish only

Table 2 Pronouns

| | A/S | P | Dative | Locative |
|-----|------------------|---------------------|---------------------|------------------------|
| 1sg | <i>ngaya</i> | <i>nganha</i> | <i>ngay</i> | <i>nganunda</i> |
| 1dl | <i>ngali</i> | <i>ngalinya</i> | <i>ngalingu</i> | <i>ngalingunda</i> |
| 1pl | <i>ngiyani</i> | <i>ngiyaninya</i> | <i>ngiyaningu</i> | <i>ngiyaningunda</i> |
| 2sg | <i>nginda</i> | <i>nginunha</i> | <i>nginu</i> | <i>nginunda</i> |
| 2dl | <i>ngindaali</i> | <i>ngindaalinya</i> | <i>ngindaalingu</i> | <i>ngindaalingunda</i> |
| 2pl | <i>ngindaay</i> | <i>ngindaaynya</i> | <i>ngindaayngu</i> | <i>ngindaayngunda</i> |

Table 3 Verb conjugations

| | Conjugation 1 | Conjugation 2 | Conjugation 3 | Conjugation 4 |
|------------------|------------------------|------------------|------------------|---------------|
| Future | <i>-li</i> | <i>-y</i> | <i>-gi</i> | <i>-rri</i> |
| Non-future | <i>-(a)y</i> | <i>-nhi</i> | <i>-nhi</i> | <i>-nhi</i> |
| Imperative | <i>-la</i> | <i>-ya</i> | <i>-nga</i> | <i>-na</i> |
| Relative present | <i>-ldaay ~ -ndaay</i> | <i>-ngindaay</i> | <i>-ngindaay</i> | <i>-ndaay</i> |
| Relative future | <i>-ligu</i> | <i>-ygu</i> | <i>-gigu</i> | <i>-rrigu</i> |

between relative future tense (purposive) or relative present tense (with adverbial or adnominal interpretations, depending on context). There are no cross-clausal coreference restrictions (such as switch reference or syntactic ergativity). Examples from Yuwaalaraay (Williams, 1980: 117–122) are shown in (4), (5), and (6).

- (4) *Girr ngaya nhama baa-nhi*
 affirm 1sgnom that jump-past
nhaadhiyaan-di, nginda garra-ldaay
 log-ablat 2sgnom cut-relpres
 'I jumped off the log that you cut'
- (5) *Nginda ngaaluurr burrulaa bayama-ndaay,*
 2sgnom fish many catch-relpres
ngay bulaarr wuu-na
 1sgdat two give-imper
 'If you catch many fish give me two'
- (6) *Ngaya yana-y walaay-gu,*
 1sgnom go-nonfut camp-dat
dhingaa dha-ligu
 meat eat-relfut
 'I am going to camp to eat meat'

Particles in Gamilaraay have scope over the whole clause and encode such semantic concepts as polar question, as in (1); affirmation, as in (4); and negation. There are different particles for negative imperative, as in (2), and negative statement, as in (7).

- (7) *Gamil ngaya gamilaraay guwaa-li*
 not 1sgnom Gamilaraay speak-fut
 'I will not speak Gamilaraay'

Bibliography

Ash A, Giacón J & Lissarrague A (2003). *Gamilaraay-Yuwaalaraay-Yuwaalayaay dictionary*. Alice Springs: Institute for Aboriginal Development.

- Austin P (1992). *A dictionary of Gamilaraay, northern New South Wales*. Melbourne: La Trobe University.
- Austin P (1993). *A reference dictionary of Gamilaraay, northern New South Wales*. Melbourne: La Trobe University.
- Austin P (1997). 'Proto-central New South Wales.' In Tryon D & Walsh M (eds.) *Boundary rider: essays in honour of Geoffrey O'Grady*. Canberra: Pacific Linguistics. 21–49.
- Austin P & Tindale N B (1986). 'Emu and brolga, a Kamilaroi myth.' *Aboriginal History* 9, 8–21.
- Austin P, Williams C & Wurm S A (1980). 'The linguistic situation in north central New South Wales.' In Rigsby B & Sutton P (eds.) *Papers in Australian linguistics, no. 13: contributions of Australian linguistics*. Canberra: Pacific Linguistics. 167–180.
- Donaldson T (1980). *Ngiyampaa: the language of the Wangaaypuwan*. Cambridge: Cambridge University Press.
- Giacon J (1999). *Gamilaraay language lessons*. Walgett: TAFE College.
- Mathews R H (1903). 'Languages of the Kamilaroi and other aboriginal tribes of New South Wales.' *Journal of the Royal Anthropological Institute* 33, 259–283.
- Mitchell T (1839). *Three expeditions into eastern Australia*. Sydney: Government Printer.
- Ridley W (1856). *Gurre Kamilaroi or 'Kamilaroi sayings.'* Sydney: Government Printer.
- Ridley W (1875). *Kamilaroi and other Australian languages*. Sydney: Government Printer.
- Williams C (1980). *A grammar of Yuwaalaraay*. Canberra: Pacific Linguistics.

Relevant Website

<http://coombs.anu.edu.au> – kamilaroi/Gamilaraay Dictionary (Coombsweb at the Australian National University; type 'Gamilaraay dictionary' into the search box).

Gə'əz

S Weninger, Philipps-Universität, Marburg, Germany

© 2006 Elsevier Ltd. All rights reserved.

Gə'əz (Gə'əz, Classical Ethiopic, Old Ethiopic) is the classical language of Ethiopia. Gə'əz belongs to the Northern branch of Ethio-Semitic, which is part of South Semitic, a subgroup of the Semitic languages. The earliest extant texts are inscriptions from the 3rd century found in Northern Ethiopia, especially in the city of Aksum. From this time, Gə'əz maintained its exclusive status as a medium of

formal communication, until the 19th century, when it gave way to Amharic. Within the Ethiopian Orthodox Church it still holds a position comparable to the status of Latin in Catholicism. Gə'əz was also the liturgical language of the Jewish community of Ethiopia, the Betä 'Īsra'el.

Gə'əz is written in a quasi-syllabic script, developed on the basis of the Sabaean alphabet. Each sign represents either a combination of a consonant and a vowel or a simple consonant, e.g., አቡነ፡ ዘበሰማያት, representing the sequence 'a-bu-nä zä-bä-sä-ma-ya-t, pronounced 'abunä zäbäsamayāt ('Our father who

art in heaven'). In its classical form, Gə'əz phonology comprises 26 consonants plus a set of four labialized velar obstruents and seven vowels. Noteworthy are sets of lateral, pharyngeal, and laryngeal obstruents as well as a set of ejectives. In the traditional pronunciation as practiced by learned Ethiopians, many consonantal distinctions are lost under the influence of modern Ethio-Semitic languages.

Gə'əz has a rich morphology based on the common Semitic system of three or four consonantal roots and vocalic patterns, with prefixes and suffixes used both in derivation and inflection, as in *šāḥafä* 'he wrote', *šāḥafku* 'I wrote', *nəṣəḥəf* 'we shall write', *šāḥafi* 'scribe', and *mäṣḥaf* 'book'. Grammatical categories of the noun are gender (male/female), number (singular/plural), and case (nominative-genitive/accusative). The verb has five tense-aspect-mood (TAM) categories (perfect, imperfect, jussive, imperative, and converb). Other categories of the verb are person, number (singular/plural), and gender (male/female, also for second person!). The syntax is based on a rather loose verb-subject-object word order, and agreement is not strict.

The vocabulary of Gə'əz is to a large degree common Semitic. However, loanwords from Cushitic languages, especially Central Cushitic languages,

form a sizable part of the Gə'əz lexicon. Contacts with the Church of Alexandria paved the way during late antiquity for many Greek and, during the Middle Ages, Arabic loanwords. On the one hand, Gə'əz shares many typological features with the Asian Semitic languages, such as Arabic or Hebrew. On the other hand, it shows features known from the modern Ethio-Semitic languages, which belong to the Ethiopian language area, also shared by the Cushitic languages, with their subject-object-verb syntax. The converb, already present in Gə'əz, is typical for this area.

Bibliography

- Dillmann A (1907). *Ethiopic grammar* (2nd edn.). [enlarged, improved, transl. edn. by Bezold C, with additions by Crichton J A]. London: Williams and Norgate. [Reprinted. Amsterdam: Philo.]
- Leslau W (1987). *Comparative dictionary of Gə'əz (Classical Ethiopic): Gə'əz-English/English-Gə'əz with an index of the Semitic roots*. Wiesbaden: Harrassowitz.
- Weninger S (2001). 'Vom Altäthiopischen zu den neuäthiopischen Sprachen.' In Haspelmath M *et al.* (eds.) *Language typology and language universals*, vol. 2. Berlin: Mouton. 1762–1774.

Georgian

B G Hewitt, SOAS, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Georgian is spoken in the Republic of Georgia, the Zakatala district of Azerbaijan, the historical Georgian regions of northeastern Turkey, by descendants of Georgians transplanted to Fereydan in Iran by Shah Abbas (17th century) and by émigré communities established in such countries as France following sovietization but growing in Russia and beyond since the collapse of the USSR in 1991.

The Georgian language (*kartuli ena*) belongs to the Kartvelian (South Caucasian) family (see **Caucasian Languages**). Georgian possesses a number of dialects, which can differ sharply from both one another (e.g., western Gurian versus northeastern Khevsurian) and the literary standard. The latter is in some respects still in the process of regularization but is based on the central Kartlian dialect, in which region lies the capital, Tbilisi (formerly T'pilisi).

The language is customarily periodized as Old Georgian (5th–11th centuries) → Mediæval

(12th–18th centuries) → Modern (post-1800). Iranian, more recently Russian, and now English lexical influences are marked; Greek, Armenian, Arabic, and Turkish loans have also penetrated.

The oldest inscription dates from circa 430 AD at a site near Bethlehem. Within Georgia, the church at Bolnisi boasts an inscription dated to 494. Iak'ob Tsurtaveli's "Martyrdom of Shushanik," apparently composed between 476 and 483, represents the first work of native literature, while the oldest dated manuscript (the Sinai Polycephalon) hails from 864. The earliest survivals exhibit peculiarities in the marking of the third person indirect object/second person subject, from which they are styled *xanmet'i* 'with extra x' or *haemet'i* 'with extra h'; the nature of this distinction (diachronic versus dialectal) has been hotly debated. Little seems to have been written during the centuries of Mongol and Tatar depredations.

Georgian is written in a unique, wholly phonemic alphabet with 33 characters from left to right without any upper- versus lower-case distinction. The modern script *mxedruli* 'military; secular' evolved in the

11th century from its precursor *k'utxovani* 'angular.' This in turn developed in the 9th century from the oldest variant *mrg(v)lovani* 'rounded,' which was probably devised in the 4th century AD on the model of Greek to facilitate the spread of Christianity, adopted as the official religion by King Mirian of Kartli during the 330s. Even after the 11th century, religious texts continued to be written in a combination of the two earliest scripts, called *xutsuri* 'ecclesiastical,' such that the oldest served as the majuscule (*asomtavruli*) to the minuscule (*nusxuri*) of its successor.

Modern Georgian has 28 consonants plus five vowels (Tables 1 and 2).

Old Georgian additionally had the voiceless uvular plosive /q/, which in standard Georgian has merged with the voiceless back fricative, plus the palatal glide /j/, sounds retained in Svan. Circumfixes abound. Verbs divide into transitives, intransitives, 'medials,' indirects (with logical subject in the dative), and a small stative class; 'medials' often appear intransitive but largely behave morphosyntactically like transitives because the relevant forms are borrowed from corresponding transitive paradigms. Georgian (with Svan) preserves the feature, assumed to have characterized Proto-Kartvelian, whereby a transitive verb's arguments are case-marked in one of three ways determined by which of three tense-mood-aspect (or *screeve*) series the given form displays. Verbs can agree with up to three arguments by virtue of the presence of two sets (A and B) of pronominal agreement affixes. The patterns of morphosyntactic behavior (with subscripts indicating affixal agreement) are (Table 3) distributed as follows (where the transitives and 'medials' are combined as Type I verbs, while intransitives, indirects, and statives are subsumed under Type II) (Table 4).

Table 1 Consonantal system

| | | | | | | | | | |
|---|---|----|----|----|-----|-----------|---|---|---|
| b | p | p' | | | | | | | m |
| | | | | | | v [v/f/w] | | | |
| d | t | t' | | | | | | n | |
| | | | dʒ | fʂ | fʂ' | z | s | l | r |
| | | | dʒ | fʃ | fʃ' | ʒ | ʃ | | |
| g | k | k' | | | | | | | |
| | | q' | | | | κ | χ | | |
| | | | | | | | h | | |

Table 2 Vowel system

| | | | | |
|---|---|---|---|---|
| i | | | | u |
| | ε | | ɔ | |
| | | α | | |

The examples below demonstrate that, while case marking in Series II follows ergative alignment, affixal agreement is accusative, creating a split-ergative configuration. Although the Series III pattern might appear to be a better candidate for illustrating ergativity, this inverted construction developed relatively late across Kartvelian out of a past (essentially intransitive) resultative. No unique Ergative morph can be reconstructed for proto-Kartvelian. The system is illustrated by the transitive 'The shepherd (a) will toss_β, (b) tossed_α, (c) (has) apparently tossed_γ, food down for his flock' versus the intransitive 'The priest (d) will drown_β, (e) drowned_β, (f) (has) apparently drowned_{β'}, where (a/d) represent Series I, (b/e) Series II, (c/f) Series III:

- (a) mʃs'q'ε ms-i sa.mʃs'q's.ɔ-s sa.ʃʃ'm. εl-s
shepherd-Nom_A *flock-Dat_B* *food-Dat_B*
 da-θ-θ-u-q'r-i-s
Prev-it_B-it_B-OV-toss-TS-he.Fut_A
- (b) mʃs'q'εms-ma sa.mʃs'q's.ɔ-s sa.ʃʃ'm.εl-i
shepherd-Erg_A *flock-Dat_B* *food-Nom_B*
 da-θ-θ-u-q'ar-a
Prev-it_B-it_B-OV-toss-he.Aor_A
- (c) mʃs'q'ε ms-s sa.mʃs'q's.ɔ-s-tvis
shepherd-Dat_B *flock-Gen-for*
 sa.ʃʃ'm.εl-i da-θ-u-q'r-i-a
food-Nom_A *Prev-he_B-OV-toss-Perf-it_A*
- (d) m.κvd.εl-i da-i-χrfʃ-ɔb-a
priest-Nom_A *Prev-Pref-drown-TS-he.Fut_A*
- (e) m.κvd.εl-i da-i-χrfʃ-ɔ
priest-Nom_A *Prev-Pref-drown-he.Aor_A*
- (f) m.κvd.εl-i da-m-χrfʃv-al-a
priest-Nom_A *Prev-Pref-drown-Suff-he.Perf_A*

The languages to have undergone most Georgian influence are naturally its congeners, Mingrelian and Svan, plus other Transcaucasian neighbors, notably

Table 3 Patterns of case marking and verb agreement

| Grammatical role | A/S | O(/P) | IO |
|------------------|------------------|------------------|---------------------|
| Pattern α | ERG _A | NOM _B | DAT _B |
| Pattern β | NOM _A | DAT _B | DAT _B |
| Pattern γ | DAT _B | NOM _A | GEN + /-tvis/ 'for' |

Table 4 Correlation of agreement pattern, verb type, and verb series

| | Series I | Series II | Series III |
|---------|----------|-----------|------------|
| Type I | β | α | γ |
| Type II | β | β | β |

Bats, Iranian Ossetic, Indo-European Armenian, and Northwest Caucasian Abkhaz. As a feudal power throughout the Caucasus and source for the spread of Christianity to the north Caucasus before the coming of Islam, Georgian has left some lexical traces here, too.

Bibliography

- Boeder W (1968). 'Über die Versionen des georgischen Verbs.' *Folia Linguistica* 2, 82–152.
- Hewitt B G (1995a). *A Georgian reader (with texts, translation and vocabulary)*. London: SOAS.
- Hewitt B G (1995b). 'Georgian: ergative, active, or what?' In Bennett D C, Bynon T & Hewitt B G (eds.) *Subject, voice and ergativity*. London: SOAS. 202–217.
- Hewitt B G (1996a). *Georgian: a learner's grammar* [2nd revised and expanded edition, 2005]. London: Routledge.
- Hewitt B G (1996b). *Georgian. A structural reference grammar*. Amsterdam: John Benjamin.
- Holisky D A (1981). *Aspect and Georgian medial verbs*. Delmar, NY: Caravan Books.
- Rayfield D (1994). *The literature of Georgia. A history*. (2nd revised edition, 2000). Richmond: Curzon Press.
- Rayfield D, Apridonidze S & Broers L (2006). *Georgian-English Dictionary*. Queen Mary College, London University: Garnett Press.
- Shanidze A (1976). *Dzveli kartuli enis gramat'ik'a* [Grammar of the old Georgian language]. German-language edition (1982). Tbilisi: University Press.
- Tschenkéli K (1958). *Einführung in die georgische Sprache* (2 vols). Zürich: Amirani Verlag.
- Tsurt'aveli I (1983). *The martyrdom of Shushanik* (8-language edition). Tbilisi: Khelovneba.
- Vamling K (1989). *Complementation in Georgian*. Lund: University Press.
- Vogt H (1971). *Grammaire de la langue géorgienne*. Oslo: Universitetsforlaget.

German

M Durrell, University of Manchester, Manchester, UK

© 2006 Elsevier Ltd. All rights reserved.

In terms of speaker numbers, German is the most important language in western and central Europe. With English, Dutch, and Frisian (Western Frisian; Northern Frisian), it belongs to the western group of the Germanic languages. In its standardized form, however, it is linguistically more conservative, having retained more of the synthetic morphology of the common ancestor. Its dialects exhibit immense variation, with a low degree of mutual intelligibility, and the standard form emerged relatively late as a consequence of the political fragmentation of central Europe.

The Speakers of German

With approximately 100 million speakers, German ranks 10th among the languages of the world, and it is the most widely spoken language in the European Union in terms of first-language speakers. Within Europe, it is exceptional in terms of the number of countries in which it is spoken. The largest proportion is in the Federal Republic of Germany (80 million), followed by Austria (7.5 million), German-speaking Switzerland with Liechtenstein (4.5 million), and Luxembourg (350 000). German also has some official status as a regional language in Belgium, Denmark, Hungary, Italy, and Romania,

and there are significant German-speaking minorities in the Czech Republic, France, Poland, Slovakia, and a number of the successor countries to the Soviet Union, notably Kazakhstan. However, where German-speaking minorities lack official status, the number of speakers is generally in decline.

Outside Europe, there are over a million active users of German in the United States and significant numbers in Argentina, Australia, Brazil, Canada, Namibia, South Africa, and a few other South and Central American countries. These numbers are now declining quite rapidly as German speakers assimilate to the majority linguistic community.

German has a few significant offshoots. The only example of a German-based creole was in Rabaul (Unserdeutsch, spoken in New Britain), but this is now vestigial. Pennsylvania Dutch (Pennsylvanish), deriving from Palatinate dialects, is spoken in parts of Pennsylvania, Ohio, and Ontario, although it is in decline outside closed religious communities such as the Amish. Since the Second World War, the local dialects of Luxembourg have no longer come to be perceived by their speakers as forms of German, and a standard Luxembourgish (Luxembourgeois) is becoming established. The most important language to have developed from German, though, is Yiddish, whose origin was the medieval German spoken in Jewish communities in central and southern Germany but that has subsequently developed into a distinct language, with a syntax and vocabulary unlike that of any variety of German.

The History of German

Modern Standard German (customarily referred to as *Hochdeutsch*, or High German) arose from the West Germanic dialects spoken by a number of peoples (Franks, Alemannians, and Bavarians) who settled between the north German plains and the Alps in the first centuries A.D. These dialects came to differ from other forms of West Germanic, and especially from the Low German dialects of the northern plains, through the consonant changes that are known collectively as the Second (or High German) Sound Shift. Because of this change, inherited voiceless plosives became affricates or fricatives (depending on whether they were word-initial or not), as illustrated by the pairs of cognates from English (which retained the original West Germanic consonants) and modern German in **Table 1**.

There are continuous written records in varieties of High German from the second half of the 8th century, and the history of the German language is usually divided into four major periods. Relatively few texts, mainly of a religious character, have survived from the first period (750–1050), known as Old High German, because Latin was the dominant language of literacy. The Middle High German period (1050–1350) saw some development of secular writing, notably in the form of chivalric verse on French models. The linguistic territory was extended during this period as High German settlers moved east across the Elbe.

The third period, Early New High German (1350–1650), saw the beginnings of a slow process of standardization. Up to this time, numerous regional varieties had been used in writing, but, with the invention of printing, certain variants started to be used more widely than in their region of origin. This development was aided by the Reformation and the prestige of Luther's writings, especially his Bible translation, which often proved decisive in the preference for a particular variant over competing forms. In this way, the process of selection, as the first stage in the process of standardization, had been largely completed by the mid-17th century, and a relatively uniform High German was being used for writing in central Germany and, significantly, also in the north, where it supplanted the native Low German.

Table 1 Second sound shift: English and German cognates

| English | German |
|---------|-------------------|
| pepper | Pfeffer |
| tin | Zinn (<z> = [ts]) |
| water | Wasser |
| book | Buch (<ch> = [x]) |

The process continued into the fourth, New High German period (1650 on) as this written variety was codified in terms of grammar, orthography, and lexis. Crucially, in the course of the 18th century, this variety gained acceptance in Catholic south Germany, Switzerland, and Austria, which had resisted a variety associated with Lutheranism and retained regionally based norms. Nevertheless, *Hochdeutsch* long remained a primarily written variety, with local dialect being the norm in speech. By the middle of the 19th century, however, a prestige pronunciation had arisen based on the north German tradition of speaking written High German precisely as it was spelled. This was formally codified for use on the stage in 1898, and it has subsequently been adopted by a majority of speakers as a spoken supraregional norm.

The Structure of Modern Standard German

Sounds and Spelling

Among the noteworthy features of German phonology is the existence of a set of affricates, specifically a labial /pf/ and dental /ts/, which have (controversially) been analyzed as single phonological units. There is also a set of front rounded vowels, as in *fühlen* [fy:lən] 'feel,' *böse* [bø:zə] 'evil.' The opposition of voicing in obstruents is neutralized in syllable-final position, as in *ich sage* [za:gə] 'I say' but *ich sagte* [za:ktə] 'I said.'

A striking feature of German orthography is the capitalization of nouns. Otherwise, it is broadly phonemic, if not precisely so. In particular, the tense/lax distinction in vowels is not marked in a uniform way, and there is a general principle that root morphemes retain a consistent spelling wherever possible. The Latin alphabet is used, with some modifications, specifically the symbol <ß>, used for /s/ in some words, and the umlaut symbol, which assists in maintaining the consistent orthographic shape of roots. The spelling of German was reformed in 1996 (after the initial codification in 1902) with the aim of eliminating inconsistencies. This reform has generated considerable controversy that is still ongoing, but the changes are relatively slight and chiefly affect the use of the symbol <ß> and punctuation.

Morphology

German is the only West Germanic language to retain extensive inflectional marking of grammatical categories, although the nature of this marking has often changed. The four noun cases, for example, are now marked primarily through the inflection of the determiner and/or the adjective, as in *der Mann* 'the man':

NOM: der Mann

ACC: den Mann

GEN: des Mannes

DAT: dem Mann

The three genders of the noun can only be identified through agreement and do not correlate consistently with any phonological or semantic features of the noun itself, for example, *der Band* ‘the volume’ (masculine), *die Hand* ‘the hand’ (feminine), *das Land* ‘the country’ (neuter).

German makes extensive use of vowel changes in inflection. Ablaut in the strong verbs (e.g., *singen* ‘sing,’ past *sang*, past participle *gesungen*) is found in other Germanic languages, but German has also morphologized the vowel fronting alternations known as umlaut in several functions, for example, in noun plurals (*der Bruder* ‘the brother,’ *die Brüder* ‘the brothers’), the subjunctive (past indicative *ich war* ‘I was,’ past subjunctive *ich wäre*), and adjective comparison (*groß* ‘big,’ *größer* ‘bigger’).

Syntax

A striking aspect of German syntax is the position of the finite verb. In main clause statements it is the second constituent, but in questions it occurs first and in subordinate clauses it occurs finally. In verb-second and verb-initial clause types, any non-finite components of the verb phrase are placed finally, forming what German linguists term a ‘bracket construction’ enclosing the other constituents:

| | | | | |
|------------------|----------------|-------------|-------------|------------------|
| deinen | Bruder | habe | ich | zufällig |
| <i>your</i> | <i>brother</i> | <i>have</i> | <i>I</i> | <i>by chance</i> |
| gestern | in | der | Stadt | gesehen |
| <i>yesterday</i> | <i>in</i> | <i>the</i> | <i>city</i> | <i>seen</i> |

‘I have by chance seen your brother in the city yesterday’

The preverbal position constitutes a topic slot that can be filled by any appropriate constituent. The position of other clause-level constituents depends on communicative criteria. The syntax of German has thus often been considered to be characteristically flat (Kathol, 2001), although a more conventional view is that it is an underlyingly SOV language, with the finite verb being moved into second position.

Regional and Social Variation in German

The German speech area exhibits much geographical variation, with significant linguistic differences among distant regional dialects. The area can be seen as a dialect continuum from the Alps to the coast, with the most important division being between the High German dialects of the center and south, which

participated in the Second Sound Shift, and the Low German dialects of the north, which did not. The Low German dialects are closer to standard Dutch than standard German, but these areas adopted standard (High) German as their language of literacy in the 17th century. The major dialect groups within High German are West Central German (in the Rhinelands and Hesse), East Central German (in Thuringia and Saxony), Bavarian (including most of Austria), and Alemannic (in the southwest, including German-speaking Switzerland).

These dialects are very diverse, and the degree of mutual comprehensibility between even geographically quite close dialects can be remarkably low. Not all Swiss Germans, who are all competent in their local dialect, can understand the dialects of the remoter Alpine valleys. Few linguistic criteria, with the possible exception of the verb-second constraint, link all the speech varieties within this continuum exclusively, and only the long-established notion that these are all, in some way, forms of German provides a connection, as does the use of the common standard, which is important for supraregional communication and as a unifying symbol of ethnic identity. The significance of the latter has become clear again following reunification in 1990. Nevertheless, there is significant variation in the codification of the standard between the various German-speaking countries, and some regional diversity is accepted in the standard even within Germany, particularly at the level of lexis (e.g., northern *Sonnabend* and southern *Samstag* ‘Saturday’).

Remarkable, too, is the sociolinguistic diversity of the German speech area in the variety of the relationships between standard German, the dialects, and other languages. German-speaking Switzerland is a classic instance of diglossia, although the nature of this diglossia is changing rapidly. South Tyrol and eastern Belgium exhibit relatively stable – and nowadays institutionalized – bilingualism, whereas in Alsace-Lorraine the centuries-old French–German bilingualism is breaking down with the rapid decline of the local dialects and the decoupling of the link between them and standard German. Within Germany and Austria the relationship between *Hochdeutsch* and the dialects varies markedly over the whole area, with the relative prestige and use of dialect increasing the further south one travels. A common pattern in central and southern Germany is the existence of a continuum of variants between near-dialect and near-standard, with speakers employing two focused varieties along this continuum that they perceive as (and label) *Hochdeutsch* and dialect. These are then used in accordance with the perceived formality of a given speech situation.

Bibliography

- Ammon U (1995). *Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten*. Berlin and New York: de Gruyter.
- Barbour S & Stevenson P (1990). *Variation in German. a critical approach to German sociolinguistics*. Cambridge, UK: Cambridge University Press.
- Besch W, Betten A, Reichmann O *et al.* (eds.) (1998–2000). *Sprachgeschichte, Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung* (2nd edn.) (3 vols). Berlin and New York: de Gruyter.
- Durrell M (2002). *Hammer's German grammar and usage* (4th edn.). London: Arnold.

- Eisenberg P (1998–1999). *Grundriß der deutschen Grammatik* (2 vols). Stuttgart and Weimar: Metzler.
- Kathol A (2001). 'Positional effects in a monostratal grammar of German.' *Journal of Linguistics* 37, 35–66.
- Stevenson P (2002). *Language and German disunity: a sociolinguistic history of East and West in Germany 1945–2000*. Oxford: Oxford University Press.
- Wells C J (1985). *German: a linguistic history to 1945*. Oxford: Clarendon Press.
- Wiese R (2000). *The phonology of German* (2nd edn.). Oxford: Clarendon Press.

Germanic Languages

M Durrell, University of Manchester, Manchester, UK

© 2006 Elsevier Ltd. All rights reserved.

The Germanic languages include Dutch, English, German, and the Scandinavian languages. From their original home in northwest Europe, they have become one of the most widely distributed language groups, being spoken on five continents by at least 550 million native speakers. By far the largest proportion of these (over 70%) are first-language speakers of English.

The Germanic Language Group

The Germanic languages constitute a distinct branch of the Indo-European language family. Since the 19th century it has been conventional to distinguish three major subgroupings within Germanic:

- *West Germanic*, including English (with many varieties and descendants, notably Scots and numerous creoles and pidgins), German (with several relatively distinct varieties and offshoots, including Low German, Luxembourgish (Luxembourgeois), Pennsylvanian (Pennsylvanish), and Yiddish), Dutch (with its descendent Afrikaans), and Frisian (with three mutually unintelligible varieties: Western Frisian, East Frisian, and Northern Frisian).
- *North Germanic* in Scandinavia, comprising a western group that includes Icelandic, the closely related Faroese (together with the now-extinct dialects of other early Norse settlements), and Norwegian (with its two codified standard varieties Bokmål and Nynorsk); and an eastern group including Danish and Swedish. The mainland Scandinavian languages are mutually comprehensible to a significant degree.

- *East Germanic*, which is usually taken to include Gothic, Burgundian, and Vandalic and possibly some further languages such as Gepidic and Rugian. However, all these languages are now extinct and the evidence that they actually constituted a discrete group with common linguistic features is slight. Little is known of them except for Gothic, for which part of a 4th century Bible translation has survived. This is the earliest continuous text in any Germanic language.

The Origins and Early History of the Germanic Languages

The genetic relationship among the Germanic languages is clear from many lexical cognates (cf. English *house*, *red*, *I gave*; Dutch *huis*, *rood*, *ik gaf*; Norwegian (Bokmål) *hus*, *rød*, *jeg gav*; Gothic *hus*, *rauþs*, *ik gaf*), and this relationship was already recognized by scholars in the 16th century. The ancestor language, usually called Proto-Germanic, was probably spoken around the North Sea and the Baltic in the first millennium B.C. Within the Indo-European family, Proto-Germanic appears to be most closely related to Italic (and possibly Illyrian and Venetic), although there are significant affinities, particularly in morphology and lexis, with Baltic (and possibly Slavonic). Nevertheless, almost one-third of the core vocabulary that can be reconstructed for Proto-Germanic lacks any Indo-European cognates, notably words relating to the sea and seafaring, such as English *boat*, *sea*, and *ship* and the words for the points of the compass. Early lexical loans from Celtic (e.g., the ancestor of German *Reich* 'kingdom') testify to the earliest contacts of the Germanic peoples, as do a significant number of lexical loans

from Proto-Germanic into Finnish (e.g., Finn. *kuningas* ‘king’). These are particularly interesting because they still retain features and forms that have disappeared in even the earliest attested Germanic dialects and can only be reconstructed for the common ancestor.

The first longer account of the Germanic peoples is by the Roman historian Tacitus in his *Germania* of 98 A.D., by which time some Germanic tribes had moved southward from the shores of the Baltic into central Europe and come into contact and conflict with the Roman Empire as it expanded northward. The first linguistic records, from the 3rd century A.D. onward, are carved inscriptions using the Runic alphabet or *futhor* (so-called after the first six letters of the alphabet). The origins of this writing system are obscure, although it appears to be based on an early Etruscan or north Italian alphabet, and it is not used for longer texts. Runic inscriptions have been found in many parts of northern and eastern Europe, and their language is remarkably uniform, irrespective of their provenance. It is sometimes erroneously referred to as Primitive Norse or Proto-Norse, but there is no evidence that its form is specifically North Germanic.

The relationship of the early Germanic dialects to one another is a matter of controversy. The assumption of an early three-way split into East, North, and West Germanic is no longer accepted, but no consensus has replaced it. A majority view at present is that the East Germanic group, specifically Gothic, split from the common ancestor first, leaving a North-West Germanic group that has been identified with the language of the early Runic inscriptions. This divided into the North Germanic and West Germanic groups. North Germanic or Old Norse is the common ancestor of the modern Scandinavian languages; it is very close, if not identical, to the attested Old Icelandic of the sagas and the *Edda*.

West Germanic, on the other hand, as the ancestor of modern Dutch, English, Frisian, and German, appears to have been a much more diffuse grouping with few common features. It is probably best considered as a dialect continuum, although the relationship of the dialects within it and the development from the earliest stages into the modern languages and dialect groups are not clear in every respect. Within West Germanic, three main complexes of dialects have been identified: North Sea Germanic (sometimes referred to as *Ingwaëonic*, following the terminology of Tacitus), Weser-Rhine Germanic (or *Istwaëonic*), and Elbe Germanic (or *Erminonic*). Modern English and Frisian have their origin in dialects of the North Sea Germanic group. Early forms of Dutch (called Old Low Franconian) are poorly attested, but modern Dutch, like Low German (and its earliest form

Old Saxon) seems to combine features of North Sea and Weser-Rhine Germanic. The principal characteristic of (High) German within West Germanic results from the Second Sound Shift, whereby inherited voiceless plosives developed into affricates or fricatives (cf. English *pepper*, *water*, *token* with German *Pfeffer*, *Wasser*, *Zeichen*). This change is commonly assumed to have occurred between the 5th and 7th centuries A.D. in the Elbe Germanic dialects and some of the more southerly Weser-Rhine Germanic dialects, which thereby combined to form the basis for modern German.

This view of the development of the Germanic languages has recently been challenged by Vennemann (1984), who put forward a new account of the development of the Germanic languages from the earliest times. In his view, the division reflected in the Second Sound Shift goes back to an underlying division within Proto-Germanic into Low and High Germanic, with (High) German thus separated from the other dialects at a very early stage. These views have not been widely accepted, but they succeeded in reopening a still ongoing debate on some of the more intractable problems in the early history of the Germanic languages.

The Germanic peoples played a significant part in the great migrations that followed the fall of the Roman Empire, and Germanic tribes spread over wide areas of Europe, although they were subsequently assimilated into the local populations. This has left substantial traces in the form of lexical loans in many languages of southern Europe (e.g., French *garder*, Italian *guardare*, and Spanish/Portuguese *guardar* from the Germanic root **wardōn* ‘watch’). The lexical influence of Frankish, a West Germanic language, on early French, is of particular importance, and some 700 such loans can be dated from the 3rd to the 8th centuries. Proto-Germanic had only two tenses (past and nonpast), although this has been extended by periphrastic forms in all the descendent languages. Proto-Germanic also retained only one nonindicative mood, combining the functions of the I-E subjunctive and optative, although this is vestigial in all the modern languages except German and Icelandic. Characteristic of all the Germanic languages, however, is an extensive set of auxiliary verbs expressing modality (called modal auxiliaries), represented

Table 1 First Sound Shift (Grimm's Law)

| <i>Indo-European</i> | <i>Latin</i> | <i>Gothic</i> | <i>English</i> |
|----------------------|--------------|---------------|----------------|
| *trejes | trēs | þreis | three |
| *dekṃ | decem | taihun | ten |
| *dhur- | forēs | daur | door |

Table 2 Verner's Law

| Indo-European | Latin | Gothic | English |
|----------------|-------|--------|---------|
| * <i>pāter</i> | pater | fadar | father |

in English by *can*, *may*, *must*, *shall*, *will*, and so on. These verbs are typically highly irregular and syntactically deviant. Traces of the I-E synthetic passive are found only in Gothic (and there only in the present tense), and the passive is expressed through auxiliary constructions in all the modern languages. However, the Scandinavian languages have also developed a new inflectional passive deriving from the grammaticalization of the reflexive pronoun.

In the noun, the three I-E genders were retained, but these have been reduced to two (common vs. neuter) in standard Dutch and much of Scandinavian. English, Afrikaans, and some dialects of Danish, however, have lost all gender distinctions. Germanic kept four of the I-E cases, nominative, accusative, genitive, and dative (although early West Germanic has traces of an instrumental). These have been subject to further attrition in almost all the modern languages, with only Icelandic (with Faroese) and standard German retaining four cases. The other languages, in particular, no longer mark any verb arguments through inflections, except in the pronouns.

The original word order of Germanic is a matter of controversy, although majority opinion favors the assumption of underlying SOV. The sentence structure in all the extant older texts, however, is very variable, and it has been viewed as flat, or nonconfigurational. The modern languages have moved to more fixed word order, with most of them

exhibiting a characteristic verb-second (V2) structure in declarative main clauses, with the verb as the second constituent and the initial position typically being occupied by the sentence topic, which may not be the subject. English is the exception here, having moved to SVO in all clause types in the early modern period. The other West Germanic languages, with the exception of Yiddish, have a characteristic SOV structure in subordinate clauses.

Bibliography

- Antonsen E H (1975). *A concise grammar of the older runic inscriptions*. Tübingen: Niemeyer.
- Bammesberger A (1986). *Der Aufbau des germanischen Verbsystems*. Heidelberg: Winter.
- Deumert A & Vandenbussche W (eds.) (2003). *Germanic standardizations. past to present*. Amsterdam and Philadelphia: John Benjamins.
- Green D H (1998). *Language and history in the early Germanic world*. Cambridge, UK: Cambridge University Press.
- König E & van der Auwera J (1994). *The Germanic languages*. London: Routledge.
- Nielsen H F (1989). *The Germanic languages: origins and early dialectal interrelations*. London and Tuscaloosa, AL: University of Alabama Press.
- Robinson O W (1992). *Old English and its closest relatives*. London: Routledge.
- van Coetsem F & Kufner H L (1972). *Toward a grammar of Proto-Germanic*. Tübingen: Niemeyer.
- Vennemann T (1984). 'Hochgermanisch und Niedergermanisch: die Verzweigungstheorie der germanisch-deutschen Lautverschiebungen.' *Beiträge zur Geschichte der deutschen Sprache und Literatur (Tübingen)* 106, 1–45.
- Weerman F (1989). *The V2 conspiracy: a synchronic and diachronic analysis of verbal positions in Germanic languages*. Dordrecht: Foris.

Gikuyu

C Githiora, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Gĩkũyũ (alternate names: Kikuyu, Gikuyu) is spoken as a first language by about one-third of Kenya's population, or about 10 million people. The speakers are known as the Agĩkũyũ (singular Mũgĩkũyũ) and they refer to their language as Gĩgĩkũyũ. The geographical dialects of Gĩkũyũ correspond to four

administrative districts, which are Mt. Kenya (Kĩrĩnyaga), Northern (Nyĩrĩ), Central (Mũrang'a), and Southern (Kiambu), but there is minimal differentiation among them. Gĩkũyũ is classified as a Thagicu language (Guthrie's zone E51) of the Central Northern Bantu family, and ultimately the Niger-Congo superfamily. Gĩkũyũ has borrowed significantly from Maasai (Purko and Ukwavi dialects), Swahili, and English. Gĩkũyũ is one of Kenya's most thriving languages with a vibrant presence in mass media and publications. It is broadcast on a number of community radio stations, and on national radio, KBC.

Pamphlets, journals, and magazines are published regularly in this language, and books written in Gĩkũyũ are published each year in Kenya and abroad. The renowned writers Ngũgĩ wa Thiong'o and Gakaara Wanjau maintain the language's international visibility by publishing major literary works in Gĩkũyũ. A standardized orthography was first published by the United Kikuyu Language Committee (UKLC) in 1947. It was revised and updated by UKLC's successor, UUGI (Ūūrũmwe na ūkũria wa Gĩgĩkũyũ) in 1984 and 2002. Two significant studies on the language have been done by Barlow (1960) and Armstrong (1967), and two grammar sketches by Gecaga (1953) and Mugane (1997). A Kikuyu-English dictionary by Benson (1964) is still in print.

Phonology

Gĩkũyũ is a tone language, where the syllable is the tone-bearing unit. In autosegmental terms, the syllable is the licenser of tone, i.e., it assigns one tone per syllable, regardless of its length (weight). These tones, however, are not marked in standard Gĩkũyũ orthography.

Gĩkũyũ has three levels of tone: high (H), mid (M) and low (L), as illustrated in the following examples: *gĩthiitũ* (MHL) 'charm', *mũthũngũ* (MHL) 'European person', *ikuua* (LLM) 'load', *njingiri* (HMH) 'musical rattles'.

Lexical tone in Gĩkũyũ changes the meaning of words: for example, *aka* (HH) 'build' vs. *aka* (MM) 'wives'; *iria* (HHM) 'sea' vs. *iria* (MML) 'milk'; *ira* (ML) 'yesterday' vs. *ira* (LM) 'snow'. Grammatical tone changes sentence or phrase meaning: for example, *nĩmaakaga* (MMHL) (habitual tense) 'they build'; *nĩmaakaga* (HMML) (past habitual) 'they used to build'.

Syllables are always open, minimally consisting of a vowel (V) or a consonant-vowel (CV) sequence, and there are heavy and light syllables. Examples: V o 'they'; VV *ooki* 'newcomers, immigrants'; CVV *tĩĩrĩ* 'soil'; CV *ke* 'take', *ma* 'truth', *ha?* 'where?'; NCwV *ngwa* 'thunder'; NCyV *ndya* 'feast'; NCV *nda* 'stomach'.

Vowels

Gĩkũyũ has seven vowels: a, e, i, ī, o, u, ū. There is contrastive vowel length so each vowel can be either long or short: for example, *tata* 'drip'/*taata* 'aunt'; *kana* 'or'/*kaana* 'child'; *kara* 'etch'/*kaara* 'little finger'. At the phonemic level, the two high vowels ī and ū are more centralized than the cardinals, i and u respectively. Examples: *ira* 'yesterday'/*ĩra* 'tell'; *uga* 'say'/*ũra* 'run away'. Derived vowels, both long and short, are numerous in the language.

The most common ones result from assimilation across morpheme boundaries, and reduplication of certain stems also induces V-length. Examples: *Githiora ūkĩra* → *Gĩthiorookĩra* 'Githiora, wake up'; *hiti igĩrĩ* → *hitiigĩrĩ* 'two hyenas'; *anake erĩ* → *anakeerĩ* 'two warriors'; *te* 'discard' → *teeateea* 'waste'; *rĩa* 'eat' → *rĩarĩĩa* 'pick at food'; *aria* 'talk' → *araaria* 'talk a bit'.

Diphthongs and triphthongs can also occur within stem boundaries, in normal speech, e.g., *kũina* 'sing', *kiuga* 'half calabash', or across phrase/morpheme boundaries, e.g., *kĩmwĩreoke* 'then tell him/her to come', *mbūkũikumi* 'ten rabbits', *rekeambeoime* 'let him/her come out first'.

The glides y, w result predominantly from vowel sequences across morpheme boundaries, whereby ī becomes y and ū becomes w, as the following examples illustrate: *kĩ-ũria* → *kyũria* 'question'; *ũ-othe* → *wothe* 'all'; *ũ-athani* → *wathani* 'reign'. The high cardinal vowels i and u do not produce glides.

Consonants

Gĩkũyũ's inventory is made up of the following consonants:

| | |
|-------------|------------------------------|
| nasals: | n m ny ng' |
| glides: | w y r (voiced alveolar flap) |
| stops: | t k |
| fricatives: | th c h b g |

As in most Bantu languages, prenasalized consonants, **nd**, **mb**, **ny**, **ng'**, **ng**, **nj**, occur very frequently in Gĩkũyũ. Sometimes they occur as derived segments: for example, when an N-morpheme (prefix) comes into contact with an obstruent, it must produce a nasal consonant, as in the following examples involving the 1st person singular subject prefix N-: *rega* 'refuse' → *ndega* 'refuse me'; *kora* 'find' → *ngora* 'find me'; *tũma* 'send' → *ndũma* 'send me'; *cuna* 'lick' → *njuna* 'lick me', etc.

Noun Classes

Gĩkũyũ is no exception to the chief characteristic feature of Bantu languages, namely the grouping of nouns into noun classes (see Table 1). These are based primarily on concordial properties, but there are discernible semantic relationships among members of each noun group.

The Concord System and Morphosyntax

The classification of nouns into these groups has important bearing on the language's grammar. Adjectives must agree with the number (singular or plural) and class of the head noun, somewhat like the behavior of the gender system in Romance languages.

Table 1 Gĩkũyũ noun classes

| Noun prefix singular/plural | Examples | General semantic content |
|-----------------------------|---|---|
| mũ-/a- | mũndũ , mũrĩmi , mwarimũ | human; kinships; professionals |
| mũ-/mĩ- | mũtĩ , mũbarĩki , mũtũng'ũ | trees, plants; diseases; parts of body |
| i-/ma- | ibuku , iniũrũ , igongona , iruga | inanimates; parts of the body; ceremony |
| kĩ-/i-/ci- | Kĩmbaruuhia , Gĩthweri , gĩthiitũ | ceremonial, religious objects; liquids; languages |
| o-/o- | thonjo , thigiriri , ngĩ , ndutuura , huria , njogu | birds, insects, animals |
| rũ- | rũũĩ , rũnyeki | long objects |
| ka-/tũ- | kamũndũ , gatu , kanyũmba | diminutives |
| ũ-/ũ- | ũthaka , ũcoorua , wagi ; ũcũrũ | abstracts; miscellaneous |
| ku-/o | kũgura ; gũthambia , kũiya | verbal nouns; miscellaneous |
| ha-/o; kũ-/o | haaha , kwene , kwao | locatives |

Table 2 Examples of concord in Gĩkũyũ

| Singular | Plural | Gloss (in plural) |
|------------------------------------|----------------------------------|---------------------|
| mũndũ ũyũ | andũ aya aniini nĩ | 'these short people |
| mũkuhĩ nĩ | marathooma | are reading |
| arathoma ibuku | mabuku | books' |
| kamũndũ gaaka | tũmũndũ tũũtũ | 'these (dim.) small |
| gakuhi | tũrkuhi | people are |
| karathooma | turathooma | reading books' |
| ibuku | mabuku | |
| riitho rĩake riinene | maitho maake | 'his/her big eyes |
| riirona wega | manene | are seeing very |
| mũno | maroona wega | well' |
| | muno | |
| ibuku rĩakwa | mabuku makwa | 'my old books got |
| riikuru rĩorĩire | makũrũ | lost there' |
| kũu | marorĩire kũu | |

The noun also must be mapped onto the verb phrase by use of a 'marker' or prefix whose form is determined by the noun's class. In the illustrating examples in Table 2, the head noun is underlined, and its concordial prefixes highlighted in bold.

When coordinate phrases involve head nouns that belong to different classes, it becomes necessary to resolve the clash of subject concords. Two strategies are possible in such cases. One strategy opts for use of the human subject plural marker for final agreement or concord on the verb phrase, irrespective of the order of the coordinating clauses, e.g., *mwarimũ (a-/ma-) na huria (ĩ-/i-) magĩcemanĩa* 'the teacher and the rhino met'; *huria na mwarimũ magĩcemanĩa nĩra* 'the rhino and the teacher met'. The second strategy takes advantage of the plural subject prefix of inanimate, nonhuman nouns of the *kĩ-/i-/ci-* noun group, e.g., *muti (ũ-/ĩ-) na karamu (ga-/tũ-) cikũunika* 'stick and pen broke'.

Word Order in Gĩkũyũ: Subject-Verb-Object (SVO)

The chief elements of the verb phrase or sentence occur according to the following template:

nĩ (focus) + SM + TAM + (*kĩ*) + (OM) + (RF) + stem + FV (O)

RF = reflexive marker which may alternately occupy object marker position (OM). FV stands for final vowel and O for object. The latter is actually outside the main template, but is necessary for a fully formed sentence.

For example:

| | | |
|----------------------|----------------|-----------------|
| Kamau | nĩ-ari-ngi-re | mũbiira |
| <i>Kamau</i> | FOC-hit-TAM-FV | <i>the.ball</i> |
| 'Kamau hit the ball' | | |
| S | V | O |

Verb morphology is very rich in an agglutinating language such as Gĩkũyũ. Derivation in the verb system, for example, is a highly productive process involving the use of extensions (infixes) inserted before the final vowel to create additional senses or meaning to the base verb. These are highlighted in bold in the following examples:

- ia**/-**thia** (causative) e.g., *koma* 'sleep' → *komia*/*komithia* 'make sleep'
- ira**/-**era** (applicative/prepositional) e.g., *thooma* 'read' → *-thoomera* 'read for/to'
- ura** (reversive) e.g., *-binga* 'close' → *-bingũra* 'open'
- ika**/-**eka** (stative) e.g., *hingika* 'be closed', *thoomeka* 'readable'
- ana** (reciprocal) e.g., *enda* 'like, love' → *-endana* 'love each other'
- anga** (diffusive) e.g., *ita* 'pour' → *itanga* 'pour all over'.

A verb may have more than one extension in a single sequence, e.g., *binga* 'close' → *hingũra* 'open' → *hingurithia* 'cause to open' → *hingũrithania* 'cause each other to open', etc.

Doubling the verb stem or reduplication diminishes the force of the action expressed by the verb. It also indicates repetition, e.g., *rũma* 'bite' → *rũmarũma* 'nibble/mince'; *rekia* 'let go' → *rekarekia* 'let go little

by little'. In the case of verb stems consisting of more than two syllables, only the first two are repeated, e.g., *hingūra* 'open' → *hingahingūra* 'open a little'; *tindika* 'push' → *tindatindika* 'push a little'.

For negation, the marker *-ti-* is inserted in the verb phrase, immediately after the subject marker, e.g., *tūtigaathii naake* 'we will not go with him/her'. However, the negative form in subordinate phrases is marked with *-ta-*, e.g., *tūgaathii tūtari naake* 'we shall go without him/her'.

Tense and Aspect Marking

The tense-aspect system is very complex in Gikūyū. Tense combines with aspect to produce a wide variety of temporal notions, in the order of TS-VB-AS-FV. (TS (tense); VB (verb); AS (aspect); FV (final vowel) which may differ according to aspect of the verb phrase). There are 9 major tenses, whose markers are highlighted in bold below:

- ra-: present progressive, current, e.g., *nīarathooma*
- Ø/-ra-: current past (within a day), e.g., *nīathoomire*
- a-: near past, e.g., *nīathoomire*
- raa-: remote past, e.g., *nīaraathomire*
- aa-: current past/future (within a day), e.g., *nīagathooma*
- kū-: near future (within the next few days), e.g., *nīekūrithooma*
- rī-: remote future, e.g., *nīndīrithooma*
- ka-: present consecutive, e.g., *nīagaathooma*
- kī-: parallel, e.g., *nīekūgithooma*

There are three aspect markers, which are *-ag-* (stative/imperative/subjunctive), *-īt-* (perfect), and *-ir-* (completive). Twenty-seven combinations are possible in theory, but only about 20 are clearly attested in the language because some combinations are constrained by semantic considerations.

Nominal Derivation

Two broad types of noun can be distinguished in Gikūyū. The first type consisting of a basic noun and its prefix only. This type of noun is not further divisible, e.g., *irigū* 'banana', *mūcinga* 'gun', *kīrīma* 'mountain'. Like the others, these nouns can derive postpositional phrases using the locative suffix, *-inī*, e.g., *irigūinī* 'in the banana', *mūcingainī* 'on/in the gun', *kīrīmainī* 'on/at the mountain', etc.

Derived nouns in Gikūyū are easily and creatively generated by use of prefixation and/or suffixation. The former consists of such prefixes as: the diminutive prefixes *ka-/ga-/tū-*, e.g., *kamūici* 'little thief', *tūmūdū* 'little people'; the augmentative prefix *kī-*, e.g., *kīmūici* 'big thief', *kīmūdū* 'big (gigantic) person'; and the

collective class prefix *ma-*, e.g., *mamūici* 'group of big thieves'.

Deverbal nouns are distinct in that they are formed through both prefixation and suffixation using the formula *mū + verb + i*, e.g., *mūtegi* 'trapper' ← *tega* 'trap'; *mwaniki* 'dryer' ← *anika* 'put to dry'; *mūkorori* 'cougher' ← *korora* 'cough'. Verbs can also function as simple verbal nouns, e.g. *gūūka gwake ti kweka* 'his coming is not good', *kwaria kūū nī kūūru* 'that talk is bad'.

Compound nouns are many in Gikūyū, and two very significant words in the speakers' cosmology are nouns of this category, that is, *Mwene + Nyaga* 'Owner of Majesty [God]', who habitually graces *Kīrī + Nyaga* 'Mountain of Majesty' or Mt. Kenya. Other such nouns include *mwaki + nyūmba* 'house-builder' and *mutua + uhoro* 'arbitrator'. Other complex nominals involve use of the associative marker *-a*, e.g., *mwaki wa nyūmba* 'builder of houses'; *nyūmba ya toro* 'sleeping room [bedroom]'.

There are a large number of nominal derivations, creating nouns each expressing a unique meaning which is determined by the semantic argument. Such thematic roles include agent, patient, manner, result, occasion, or locative. For example, agent nominalization is very productive. It involves a nominalizing suffix *-i*, and an appropriate noun class prefix. A number of proper names of Agikūyū are thus derived, e.g., *Mwaniki* 'one who cures skins' or *Mūrīthi* 'one who herds'. The agent type of nominalization is also the source of many synthetic compounds such as *mūenda-andū* 'a person who is kind to people'. Patient nominalization involves transitive verb stems with the suffix *-o* to describe 'the act of', e.g., *mwūiro* 'telling oneself, self-deception', while reflexive *-ī-* can be added to the stem, e.g. *mwīendo* 'state of liking oneself, selfishness'. Product nominalization uses the *mū-* prefix or *kī-/gī-*, and the final vowel changes from *a* to *o*. Intransitive verbs mostly participate in this process, e.g., *gīcaambio* 'defamation' ← *caambia* 'defame'; *gīthoomo* 'education' ← *thooma* 'read'; *mwandiko* 'handwriting' ← *andika* 'write'. Location-type nominalization involves the *i-* prefix for consonant-initial stems, and *rī-* for V-initial stems, e.g., *ikuūiro* 'place of loading' ← *kuua* 'carry'; *ithūiro* 'sunset' ← *thūa* 'go down'. Nominalization of manner involves prefixation with *mū-* and suffixation with *-ire*, e.g., *mūkuūire* 'manner of carrying' ← *kuua* 'carry'; *mwandikīre* 'writing style' ← *andika* 'write'. Occasion nominalization can occur with any verb type by prefixing with *i-* and leaving the final vowel unchanged. The resulting noun refers to occasions of an event specified by the verb stem, e.g., *iceera* 'visit' ← *ceera* 'visit'; *iruga* 'feast' ← *ruga* 'cook'.

Bibliography

- Armstrong L (1967). *Phonetic and tonal structure of Gikũyũ*. London: Dawsons of Pall Mall.
- Barlow A H (1960). *Studies in Kikuyu grammar and idiom*. Edinburgh: Blackwood & Sons.

- Benson T G (1964). *Kikuyu-English dictionary*. London: Oxford University Press.
- Gecaga B M (1953). *A short Kikuyu grammar*. Nairobi: Macmillan.
- Mugane J M (1997). *A paradigmatic grammar of Gikũyũ*. Stanford, CA: CSLI.

Goidelic Languages

P Ni Chatháin, University College Dublin, Dublin, Ireland

© 1994 Elsevier Ltd. All rights reserved.

‘Goidelic’ is the term used to denote a linguistic subgroup of Celtic spoken in Ireland, Scotland, and the Isle of Man. It may describe both the original unattested predecessor of primitive Irish and its direct historical descendants, i.e., Old, Middle, and Modern Irish, Scottish Gaelic, and Manx.

The earliest corpus of Goidelic is the collection of names inscribed in Ogham cypher on stone and datable from around the fourth to the sixth centuries. These archaic Irish names reflect a state of the language comparable to Continental Celtic and Classical Latin and to some degree the intermediate stages between proto-Celtic and Old Irish (see Celtic).

Old Irish retains the IE system of nominal declension, with masculine, feminine, and neuter gender and a reduction to five cases – nominative, vocative, accusative, genitive, and dative. The old Irish noun reflects archaic formations such as the remains of the ancient hetero-clitic declension as in *arbor* (‘corn’) with gen. *arbe*, and dvandva substantival compounds as in *gaísced* (‘spear and shield’) and *fotlethet* (‘length and breadth’).

Nominal forms are the verbal adjective, the verbal of necessity, and the verbal noun. The verbal noun is used in a wide variety of constructions but it retains its nominal character in governing the genitive and not the accusative, e.g., *imgabáil uilc (gen.) do dénum* (‘to avoid doing evil’).

Conjugated prepositions, usually styled prepositional pronouns, e.g., *dom* ‘to me’ (<do ‘to’ + 1 sg. pronominal element), form part of a complex pronominal system ranging from emphatic enclitic particles, e.g., *mo chenn-sa* (‘my head’), to obsolescent suffixed object pronouns, *gaibthi* < *gaibid-i* (‘he takes it’), and the more productive infixed object pronouns, e.g., *no-t-gaib* (‘he takes you’).

Active and medio-passive are found in the Irish verb which has inherited -r endings in the dependent and passive forms. The Old Irish verb had primary stems in the present indicative, present subjunctive, future, and preterite. The first three formed secondary or past tenses which are original to Celtic. Thus:

| | Present | Subjunctive | Future | Preterite |
|-----------|------------------|-----------------|-----------------|---------------|
| Primary | <i>gaibid</i> | <i>gabaid</i> | <i>gébaid</i> | <i>gabais</i> |
| Secondary | <i>no-gaibed</i> | <i>no-gabad</i> | <i>no-gébad</i> | — |

There were two sets of endings in the primary stems, one used in independent, the other in dependent position. This opposition, originally belonging to the IE present/aurist system, was transformed in Irish into the opposition absolute/conjunct for simple verbs and deuterotonic/prototonic for compound verbs and spread through most parts of the verb, e.g., *gaibid* (‘he takes’), *ní gaib* (‘he does not take’), and *gébaid* (‘he will take’), *ní géba* (‘he will not take’). The secondary tenses in Irish had a distinctive set of endings and display no -r endings.

The preterite reflects a coalescence of the aorist and perfect of the parent language. The perfect is expressed in OI by the use of preverbs most notably *ro*, e.g., *as-beir* (‘says’), pret. *as-bert*, and perf. *as-rubart*.

The verb ‘to be’ has a two-fold division in Irish. The copula consists of proclitic forms and denotes the connection between predicate and subject e.g., *it móir ind fhir* (‘the men are big’). The substantive verb occurs with a prepositional or adverbial phrase, e.g., *attá oc techt* (‘he is going’).

It is possible that the Norse presence in Ireland, for nearly 400 years from the end of the eighth century, may have been a force for transition from the tenth century onwards associated with the middle Irish period (950–1200). Morphological simplifications characterize these changes for the most part:

- a. Deponent verb becomes obsolete, *molaithir* ('he praises') > *molaid*.
- b. Distinctions absolute/conjunct and deuterotonic/prototonic become blurred beside the actual decline of the compound verb, *do léici* ('he casts') *ní teilet* ('he does not cast'), is superseded by *teilcid* ('he casts'), a back-formation from the prototonic with the ending of the simple verb.
- c. Infix pronouns give way to independent object pronouns, *no-m-marba* ('he kills me'), becomes *marbaid mé*.
- d. Predicative adjectives are no longer declined and the inflected copula is reduced to an impersonal third sg. form, *is mór na fir sin* ('those men are big').

By the beginning of the thirteenth century Middle Irish had disappeared and a new literary Irish based on vernacular usage became the norm. Although phonetically conservative, it was grammatically of its time, with an innovative vocabulary freely borrowing from Latin, French, and English. This standard Irish remained for over 400 years as the language of professional Gaelic men of learning in Ireland, Scotland (for almost another century), and the Isle of Man.

It is described in profuse detail in grammatical tracts datable to the sixteenth century and unique for the European languages of the time. Speech parts were classified as: *focal* concrete noun, adjective, and stressed pronoun; *pearsa* verbal noun and verb; and *iarmbérla* particle, which comprised all proclitics including the article, the copula, and prepositions. The system is not that of Latin. It had been observed that the threefold division corresponds to that of Arabic grammar but no connection has been traced. The amassing of such a wealth of material with its meticulous classification and thousands of citations had as its aim teaching of verse composition in the classical standard language.

With the collapse of the Gaelic world and its aristocratic culture the subliterate dialects emerge in their divergent forms. Western Gaelic or Modern Irish in the early 1990s comprises three principal dialect areas with native Irish speakers largely confined to the western seaboard of three eponymous provinces, Ulster, Connacht and Munster. There are few monoglot speakers.

Since the Irish state was set up in 1922 the Irish language has been taught in the schools. The orthography was simplified in 1948 and a standard grammar based on the main dialects has been in official use since 1953 with subsequent revisions.

Eastern Gaelic had become Scottish Gaelic and Manx. Shared features are:

- | | |
|---|--|
| Nouns declined only for number with the commonest plural endings in -n: | (a) |
| Sc: cáirdean Mx. caarjin Ir. cáirde 'friends' | |
| Coalescence of future and present forms: | (b) |
| Ir. caillfidh 'will lose' | |
| caillidh 'loses' | Sc. caillidh, Mx. caillee, 'will lose' |
| Periphrasis used for non-habitual present tense: | (c) |
| Ir. an gcreideann tú? | |
| Sc. am bheil thu creidsinn? 'Do you believe?' | |
| Mx. vel oo credjal? | |
| Negative particle <i>ní</i> replaced by <i>cha</i> . | (d) |

The oldest document of length in a Scottish Gaelic recognizably different from Irish dates from the early sixteenth century and that in Manx from the very beginning of the seventeenth. Manx orthography differed from that of its sisters by being based on English. Scottish Gaelic is spoken in the Hebrides and the western mainland of Scotland and in Canada (Nova Scotia). Irish and Scottish Gaelic are used extensively in the mass media.

Sample Texts

Irish

Ár n-Athair, atá ar neamh: go naofar d'ainm.
/a:r nahir¹ aata: er¹ n'a:w gəni:fər dan¹ im¹
Go dtaga do ríocht.
gədaɡə də r'iəxt/
'Our Father who is in heaven: be-blessed thy name.
may-come thy kingdom'

Scottish Gaelic

Ar n-Athair a tha air nèamh; gu naomhaichear d'ainm.
Thigeadh do rìoghachd.

Manx

Ayr ain, t'ayns niau: casherick dy row dty ennym. Dy jig dty reeriaght.

All three languages were threatened by the spread of English. The last speaker of Manx died in 1974 though it is in use as a second language; some 80 000 in Scotland speak Scottish Gaelic (plus fewer than 10 000 in Nova Scotia); and, although Irish is taught to all pupils in the Republic of Ireland and censuses show returns for 1 million speakers, it is estimated that it is the native language of some 60 000.

Bibliography

- Bergin O (1938). *The Native Irish Grammarian* (Vol. xxiv). PBA.
Broderick G (1984). *A Handbook of Late Spoken Manx* (3 vols.). Tübingen.

Greene D (1972). *The Irish Language*. Cork.
 Jackson K (1951). *Common Gaelic: The Evolution of the Goidelic Languages* (vol. xxxvii). PBA.
 Ó Cuív B (ed.) (1969). *A View of the Irish Language*. Dublin.

O'Rahilly T F (1972). *Irish Dialects Past and Present*. Dublin.
 Thomson R L (1981). *Lessoonyn Sodjey 'sy Ghailck Vanninagh*. Douglas.

Gondi

P S Subrahmanyam, Annamalai University, Bangalore, India

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Gondi (also written Gonḍi with ḍ representing the actual pronunciation) is a tribal language of the Central subgroup of the Dravidian family with a population of more than two million, the largest among the tribal languages of that family. Gonds of almost all regions call themselves *kōytūr* and their language *kōyāṇ*. It is spread over the five states of Madhya Pradesh, Chattisgarh, Maharashtra, Andhra Pradesh, and Orissa in India. Further, the tribe has roughly thirty social subdivisions, the well known among them being Raj Gonds, Madias, Murias, and Koyas. Because of the vast area occupied and the social subdivisions, the language has a number of dialects, which could be considered as separate languages rather than dialects because of the great variation exhibited. There is great scope for further study on the various aspects of the language. (Unless otherwise mentioned, all examples given below are from the Adilabad dialect of Andhra Pradesh.)

Phonology

Gondi contains ten vowels like the majority of its sister languages as shown in Table 1. The core consonant system that is common to most of the dialects is presented in Table 2. However, there are additional consonants in most of them. The Adilabad dialect not only retains the aspirated stops of Marathi loans, but

the stops in some of the native words also take on aspiration as an additional feature, e.g., *phōṛd* 'sun.' In the northern area, while some dialects retain the contrast between *r* and *ṛ*, these two sounds merge into *r* in some others and into *ṛ* in still others. The dialects in the remoter parts of Chanda, Bastar (Muria Gondi), and the Koya area of Malkangiri have two *r* sounds, one a normal trill (*r*) and the other a strong trill (written *r̥*/R); the latter represents Proto-Dravidian. **t*. In the Hill-Maria dialect, the original **r* changes to a postvelar voiced fricative G; this dialect, on the other hand, preserves *ʃ* and *ɳ*. Koya shows glottalized *t'* and *k'* and retroflex *ṇ* but does not have *h*.

One important phonological feature that divides the entire Gondi area into three dialects is the development of Proto-Dravidian **c*- in the word-initial position. The dialects of the north and west (e.g., Adilabad, Betul) show *s*- for it, those of the south and east show *h*- (e.g., Chanda, Bastar, Kanker), while it is elided in the Hill-Maria and the Koya dialects, e.g., *sovvōr/hovvōr/ovōr* 'salt.'

In some of the dialects, the contrast between short and long vowels is found only in the first syllable and vowel length is neutralized in noninitial syllables. In the Adilabad dialect, all vowels in noninitial syllables are invariably long, except in some recent loans. On the other hand, all vowels in such syllables are short in the Muria dialect.

Table 1 Vowels of Gondi

| | Front | | Central | | Back | |
|------|-------|------|---------|------|-------|------|
| | Short | Long | Short | Long | Short | Long |
| High | i | ī | | | u | ū |
| Mid | e | ē | | | o | ō |
| Low | | | a | ā | | |

Table 2 Consonants of Gondi (Core System)

| | <i>L^a</i> | <i>D^a</i> | <i>R^a</i> | <i>P^a</i> | <i>Vel^a</i> | <i>G^a</i> |
|-----------|----------------------|----------------------|----------------------|----------------------|------------------------|----------------------|
| Stop | | | | | | |
| VL | p | t | t̪ | c | k | |
| VD | b | d | d̪ | j | g | |
| Nasal | m | n | | | ŋ | |
| Fricative | | s | | | | h |
| Lateral | | l | | | | |
| Trill | | r | | | | |
| Flap | | | r̥ | | | |
| Semivowel | v | | | y | | |

^aAbbreviations: D, dental; G, glottal; L, labial; P, palatal; R, retroflex; VD, voiced; Vel, velar; VL, voiceless.

Syntax

Word Classes

The following word classes may be recognized for Gondi: nouns (including pronouns and numerals), verbs, adjectives, adverbs (including expressives), particles, and interjections.

An adjective in Gondi agrees with the noun that is qualified in number and gender; this agreement rule, which is alien to Dravidian, is taken over from Indo-Aryan, for example:

| | |
|----------------------|-----------------|
| persā | marā |
| ‘big | tree’ |
| persā-ŋ | marā-k |
| <i>big-NON-M.PL.</i> | <i>tree-PL.</i> |
| ‘big trees’ | |
| pers-ōr | māynāl |
| <i>big-M.SG.</i> | <i>man</i> |
| ‘big man’ | |
| pers-ūr | māynāl-īr |
| <i>big-M.PL.</i> | <i>man-PL.</i> |
| ‘big men’ | |

Adverbs may be divided into those of (a) time (e.g., *ninnē* ‘yesterday,’ *nēnd* ‘today,’ *nārī* ‘tomorrow’), (b) place (e.g., *iggā* ‘here,’ *aggā* ‘there,’ *baggā* ‘where’), and (c) manner (e.g., *bhāy* ‘much,’ *cokkōt* ‘well’).

Examples for particles include *ānī* ‘and,’ *mattī/battī* ‘but,’ *pajjā* ‘afterwards.’

Examples for interjections include *hav* ‘yes,’ *āyō* ‘no.’

Word Order

The favored word order in Gondi is S(ubject) O(bject) V(erb), for example:

| | | | |
|---------------------|--------------|--------------|-------------------------|
| vōr | nā-kūn | kottā-ŋ | sī-t-ōr |
| <i>he</i> | <i>I-DAT</i> | <i>money</i> | <i>give-PAST-3M.SG.</i> |
| ‘He gave me money.’ | | | |

Gender and Number

Gondi shows a two-way distinction in gender, differentiating between men and all others (including women) in the singular and plural; but the masculine (+feminine) plural form also denotes, apart from men, a group of persons that contains at least one man:

| |
|---|
| vōr |
| ‘he’ |
| vūr |
| ‘those men/men(/man) and women(/woman)’ |
| ad |
| ‘she/it’ |
| av |
| ‘those women/nonhumans’ |

Finite Verb Agreement

The finite verb shows agreement with the subject pronoun (or a corresponding noun in the case of the third person) by a change in the personal suffix (see Table 4), for example:

| | |
|-----------|----------------------|
| nannā | vā-t-ōn |
| <i>I</i> | <i>come-PAST-1SG</i> |
| ‘I came.’ | |

For agreement between the demonstrative adjectives and the nouns qualified with regard to number, see ‘Word Classes.’

Noun Morphology

Most of the nominal bases are underived but a few masculine ones have the suffix *-āl* and a few feminine ones *-ār* or *-ī*, for example:

| |
|------------------|
| novr-āl |
| ‘bridegroom’ |
| novr-ī |
| ‘bride’ |
| sēl-ār |
| ‘younger sister’ |

A nominal base is followed by the plural suffix when plurality has to be expressed. A case suffix/postposition, which occurs at the end, is preceded in most cases by one of the oblique suffixes *-d-*, *-t-* or *-n-* (conditioned variants).

Plural Suffixes

Most of the masculine nouns take the plural suffix *-r* (conditioned variants *-īr* and *-ūr*), for example:

| |
|---------------|
| kāndī-r |
| <i>boy-PL</i> |
| ‘boys’ |

The plural suffix in the nonmasculine nouns has three variants: after a vowel, *-ŋ*; after a consonant, *-k*; after a disyllabic noun ending in *l*, *-īk*. There are some exceptions for the conditionings indicated, for example:

| |
|-------------------|
| paṇḍī-ŋ |
| <i>fruit-PL</i> |
| mal-k |
| <i>peacock-PL</i> |
| ‘peacocks’ |
| ḍuvvāl-īk |
| <i>tiger-PL</i> |
| ‘tigers’ |

Case Suffixes and Postpositions

The nominative is unmarked.

The accusative-dative suffix is *-ūn*. The case is unmarked in the case of inanimate nouns:

kōndā-t-ūn
bull-OBL-ACC/DAT
 ‘bull (accus)/to the bull’

The instrumental-locative suffix is *-ē*:

kay-d-ē
hand-OBL-INSTR/LOC
 ‘with/in the hand’

The ablative suffix is *-āl*:

kuhī-t-āl
well-OBL-ABL
 ‘from the well’

The genitive suffix is *-ā* (variants *-nā*, *-vā*):

kuhī-t-ā
well-OBL-GEN
 ‘of the well’

Examples for postpositions are: *+aggā* ‘in, near,’ *+karūm* ‘near,’ and *+pajjē* ‘behind’:

rō-t + pajjē
house-OBL-POSTP
 ‘behind the house.’

Pronouns

The personal pronouns are:

nannā ‘I’ marāt ‘we’
 nimmē ‘you (sg.)’ mirāt ‘you (pl.).’

The deictic and the interrogative pronouns are given below.

| | | | |
|---------|-----------|---------------|------------------|
| Distant | Proximate | Interrogative | |
| vōr | vēr | bōr | ‘he’ |
| vūr | vīr | būr | ‘they (m.[+f.])’ |
| ad | id | bad | ‘she, it’ |
| av | iv | bav | ‘they (non-m.)’ |

The reflexive pronouns are *tannā* (sg.) and *tam-mō(t/k)* (pl.).

Numerals

Dravidian cardinal numerals are preserved only up to seven in the Adilabad dialect (up to six in the Muria dialect and up to ten in the Betul dialect); the word for ‘hundred’ is also preserved in most of the dialects). These native numerals have separate forms for masculine and nonmasculine (see Table 3). The remaining numerals as well as all ordinals are borrowed in each dialect from the major language of the area. The higher numerals borrowed from Marathi

Table 3 Native numerals of Gondi

| Numerals | Nonmasculine | Masculine |
|----------|--------------|-----------|
| 1 | undī | vorō-r |
| 2 | rand | i-vvīr |
| 3 | mūd | mu-vvīr |
| 4 | nālūṅ | nāl-vīr |
| 5 | siyyūṅ | siy-vīr |
| 6 | sārūṅ | sār-vīr |
| 7 | ērūṅ | ēr-vīr |
| 100 | nūr | |

in the Adilabad dialect add the classifier *jhank* when they qualify a masculine noun and *jhanik* when they qualify a feminine noun, for example:

āṭh jhank mās-ūr
eight CLASSIF man-PL
 ‘eight men’

āṭh jhanik veylō-k
eight CLASSIF woman-PL
 ‘eight women’

Verb Morphology

Verb Bases

A verb base in Gondi can be simple or complex. The complex base is formed from the simple one by the addition of the transitive-causative suffix *-ūs* (conditioned variants: *-h*, *-pūs*). This suffix converts an intransitive into a transitive and an underived transitive into the corresponding causative, for example:

| | |
|----------------|----------------------------|
| un- ‘to drink’ | u-h- ‘to make to drink’ |
| udd- ‘to sit’ | u-ppūs- ‘to seat’ |
| aṭṭ- ‘to cook’ | aṭṭ-ūs- ‘to cause to cook’ |

Finite Verbs

A finite verb is distinguished from a nonfinite one by the presence of the personal suffix at the end of the former; a finite verb of the indicative mood is of the following structure (the past negative and the debitive are exceptions to this):

Verb Base + Tense/Negative Suffix + Personal Suffix.

There are five types of finite verb: past (suffix: *-t-*), present-future (suffixes: *-ānt-*, *-nt-*), future (suffixes: *-ak-*, *-k-*, *-n-*, *-ān-*, *-ār-*), past habitual-cum-irrealis (suffixes: *-nd-*, *-d-*) and negative (nonpast; suffixes: *-ō-*, *-v-*) (see Table 4 for all the forms of *pā-* ‘to beat’).

The imperative suffixes are: 2sg. *-ā* (variants: *-m*, *-Ø*), 2pl. *-āṭ* (variants: *-mṭ*, *-ṭ*), for example:

Table 4 Finite forms of *pā-* 'To Beat'

| | <i>Past</i> | <i>Pres.-fut.</i> | <i>Fut.</i> | <i>Past hab./irrealis</i> | <i>Neg.</i> |
|--------------|-------------|-------------------|-------------|---------------------------|-------------|
| 1 sg. | pā-t-ōn | pā-nt-ōn | pā-k-ā | pā-nd-ūn | pāy-ō-n |
| 1 pl. | pā-t-ōm | pā-nt-ōm | pā-k-ōm | pā-nd-ūm | pāy-ō-m |
| 1 pl (incl). | | | pā-k-āt | | |
| 2 sg. | pā-t-ī | pā-nt-ī | pā-k-ī | pā-nd-ī | pāy-v-ī |
| 2 pl. | pā-t-īt | pā-nt-īt | pā-k-īt | pā-nd-īt | pāy-v-īt |
| 3 msg. | pā-t-ōr | pā-nt-ōr | pā-n-ūr | pā-nd-ūr | pāy-ō-r |
| 3 m (+f) pl. | pā-t-ēr | pā-nt-ēr | pā-n-īr | pā-nd-īr | pāy-ēr |
| 3 fmsg. | pā-t-ā | pā-nt-ā | pāy-ār | pā-nd-ū | pāy-ō |
| 3 fnpl. | pā-t-āṇ | pā-nt-āṇ | pā-n-ūṇ | pā-nd-ūṇ | pāy-ō-ṇ |

aṭṭ-ā
cook-2sg
 'Cook!'

aṭṭ-āt
cook-2pl
 'Cook (pl.)!'

The corresponding negative imperative has the negative suffix *-m-/v-* between the base and the personal suffix:

aṭ-v-ā/aṭ-m-ā
cook-NEG-2sg
 'Don't cook (sg.)!'

aṭ-v-āt/aṭ-m-āt
cook-NEG-2pl
 'Don't cook (pl.)!'

The past negative has the suffix *-makī(n)*, which is invariable for person, number, and gender:

tar-makī(n)
bring-PASTNEG
 '(One) did not bring.'

The debitive has the suffix *-ānā*, for example:

tind-ānā
eat-DEB
 '(One) must eat.'

Nonfinite Verbs

The past participle has the suffix *-sī* (variants: *-cī*, *-jī*):

vā-sī
come-PASTPP
 'having come'

The present participle has the suffix *-sēr* (variants *-cēr*, *-jēr*), for example:

un-jēr
drink-PRESPP
 'while drinking'

The negative participle has the suffix *-vāk*, for example:

veh-vāk
tell-NEGPP
 'without telling'

The conditional (with the conditional suffix *-ēkē*) has three subtypes as illustrated below.

- (1) Past conditional:
 vā-t-ēkē
come-PAST-CONDI
 'when one came'
- (2) Nonpast conditional:
 vā-n-ēkē
come-NON-PAST-CONDI
 'when one is coming'
- (3) Negative conditional:
 vāy-v-ēkē
come-NEG-CONDI
 'when one does not come'

Two types of verbal nouns are commonly used. One with the suffix *-mār* simply denotes action:

aṭ-mār
cook-VN
 'cooking'

The other one with *-vāl* denotes action, agent, or goal:

veh-vāl
tell-VN
 'telling/one who tells/the matter that is told'

Bibliography

- Andres S (1977). 'A description of Muria Gondi phonology and morphology.' Ph.D. diss. Deccan College Postgraduate and Research Institute, Pune.
- Bhattacharya S (1968). 'Muria morphology.' *Bulletin of the Anthropological Survey of India* 17, 336–373.
- Burrow T & Bhattacharya S (1960). 'A comparative vocabulary of the Gondi dialects.' *Journal of the Asiatic Society (Bengal)* 2, 73–251.
- Grierson G A (1906). *Linguistic survey of India, vol. IV: Munda and Dravidian languages*. Calcutta.

- Lind A L (1913). *A manual of the Mardia language*. Kedgaon.
 Mitchell A N (1942). *A grammar of Maria Gondi as spoken by the Bison Horn or Dandami Marias of Bastar State*. Jagdalpur.
 Natarajan G V (1977). 'Adjectival concord in Gondi.' *Indian Linguistics* 38, 156–160.
 Natarajan G V (1985). *Abhujmaria grammar*. Mysore: Central Institute of Indian languages.
 Nimsarkar P D (1991). 'A linguistic analysis of the Maḍia dialect of Gaḍchiroli district.' Ph.D. diss. Nagpur University.
 Steever S B (1998). 'Gonḍi.' In Steever S B (ed.) *The Dravidian languages*. Routledge. 270–297.
 Subrahmanyam P S (1968). *A descriptive grammar of Gondi*. Annamalainagar: Annamalai University.
 Trench C G C (1919–21). *Grammar of Gondi as spoken in the Betul District, Central Provinces, India, vol. I, Grammar, vol. II, Vocabulary, folk-tales, etc.* Madras.
 Tyler S A (1969). *Koyā an outline grammar (Gommu dialect)*. Berkeley: University of California.
 Umamaheswara Rao G (1987). 'A comparative study of the Gonḍi dialects.' Ph.D. diss. Osmania University.
 Von Fueror-Haimendorf C (1979). *The Gonds of Andhra Pradesh*. London: George Allen & Unwin.
 Williamson H D (1890). *Gondi grammar and vocabulary*. London.

Gothic

J M Y Simpson, University of Aberdeen,
Aberdeen, UK

© 2006 Elsevier Ltd. All rights reserved.

Gothic is the only documented member of the East Germanic group of Germanic languages.

Early History and Wulfila's Gothic

From around the second century B.C.E. onward, various Gothic tribes migrated from southern Scandinavia to eastern and southeastern Europe, following the Vistula and Danube rivers and reaching the north of the Black Sea area by the middle of the 3rd century C.E. Tribes of Ostrogoths settled to the east of the River Dniestr and tribes of Visigoths to the west of it. In the 4th century C.E., in present-day Bulgaria, a translation of much of the Bible, based on a Greek text used in the diocese of Constantinople, was made by Bishop Wulfila (Ulfilas), a Visigoth. Portions of this (by far the greater part from the New Testament) have survived and, since this translation is the earliest literary record in any Germanic language, these are documents of outstanding importance for Germanic and Indo-European linguistic history. Significant writings in other Germanic languages begin to appear only four centuries later.

Most of Wulfila's translation has been lost. Just over half the Gospels are preserved in the surviving pages of the splendid Codex Argenteus, an Ostrogothic manuscript probably written in present-day Italy and dating from the 5th century; it is now in Uppsala. Other portions of the Gospels and of the Pauline Epistles, together with three chapters of Nehemiah, survive

in various other manuscripts, the majority of them in Milan.

Wulfila designed an alphabet clearly based on that of Greek; a version used by later scribes appears in **Figure 1** with a widely used transliteration. (The two untransliterated symbols were used only to form numerals.) ⟨q⟩ is taken to be /kw/, ⟨þ⟩ to be /θ/, and ⟨h⟩ to be /m/ or /hw/. It may be assumed that this system is phonemic, though the following digraphs have the probable values: ⟨ei⟩ = /i:/, ⟨au⟩ = /ɔ:/, ⟨ai⟩ = /ɛ/, and ⟨gg⟩ = /ŋg/. It is likely that intervocalically the letters ⟨b⟩, ⟨d⟩, and ⟨g⟩ denoted the voiced fricative allophones [β], [ð], [ɣ]. All other symbols may be given their IPA values (see **Figure 1**).

Wulfila's Gothic shows the typically Germanic features of

1. the 'First Sound Shift' development from Proto-Indo-European, thus Latin *pes* 'foot,' *tu* 'thou,' *centum* 'hundred' versus Gothic *fotus*, *þu*, *hund*;
2. strong versus weak verbs (respectively with vowel-change (*Ablaut*) versus a dental suffix in the past tense) e.g., *tiuhan* 'lead,' *haban* 'have' (infinitive) but *tauht*, *habaida* (1 sg. past);
3. weak declension of adjectives (used after a determiner) versus strong (used elsewhere), e.g., *sa goda hlaifs* 'the good bread,' *gods hlaifs* 'good bread.'

A B C D E F G H I J K L M N
 a b c d e f g h i j k l m n
 O P Q R S T U V W X Y Z
 o p q r s t u v w x y z

Figure 1 Reprinted from *Concise encyclopedia of language and religion*, Sawyer *et al.* (eds.), J. M. Y. Simpson, 'Gothic,' p. 189, Copyright (2001) with permission from Elsevier.

Among various archaic features are the retention of a masculine nominative ⟨-s⟩, e.g., *fisks* ‘fish’ and reduplicating verbs, e.g., *greten* ‘weep’ (infinitive), *gaigrot* (1 SING PAST).

Morphologically, Gothic is fairly complex. Verbs show inflections for (a) past and nonpast tenses; (b) indicative, subjunctive (sometimes called ‘optative’), and imperative moods, plus an infinitive form; (c) active and passive voices; (d) first, second, and third persons; and (e) singular, dual, and plural numbers. Nouns, adjectives, and pronouns show inflections for (a) masculine, feminine and neuter genders; (b) singular and plural numbers; and (c) nominative, accusative, genitive, and dative cases. However, there are not distinct morphological forms for every possible combination of these grammatical categories. Pronouns have in addition distinct forms for the dual number.

The syntax of Wulfilas’s text is very strongly influenced by that of his Greek source: the original word order is closely followed and many Greek devices, such as characteristic participial constructions, are imitated. As a result almost no information is available on what a native Gothic syntax might have been.

Other records of early Gothic are meager in the extreme. They include fragments of a commentary on St John’s Gospel (the *Skeireins*), a few marginal notes on Latin manuscripts (including a title deed from Ravenna, now in Naples), and one or two runic inscriptions that have been claimed to be in Gothic.

Sample Text

| | | | | | |
|--------------------|-------------------------|----------------------|--------------------------|---------------------|--------------|
| Jah þan | bid-jaiþ | ni | si-jaiþ | swaswe | þai |
| and when | pray-2.PL.PRES.SUBJ | not | be-2.PL.PRES.SUBJ | like | the-M.PL.NOM |
| liut-ans | unte | fri-jond | in | gaqumþ-im | |
| hypocrite-M.PL.NOM | for | love-3.PL.PRES.INDIC | in | synagogue-E.PL.DAT | |
| jah waiht-am | plap-jo | stand-andans | bid-jan | | |
| and | corner-M.PL.DAT | street-E.PL.GEN | stand-PRES.PART.M.PL.NOM | pray-INF | |
| ei | gaum-jaindau | mann-am. | amen | qib-a | |
| so that | see-3.PL.PRES.SUBJ.PASS | man-M.PL.DAT | truly | say-1.SG.PRES.INDIC | |
| izw-is | þatei hab-and | mizd-on | sein-a | | |
| you-DAT.PL | that have-3.PRES.INDIC | reward-ESG.ACC | their-ESG.ACC.STRONG | | |

‘And when you pray, do not be like the hypocrites, for they like to pray in synagogues and on street corners standing up, so that they may be seen by people. Truly I say to you that they have their reward.’ (Matthew vi:5)

Other East Germanic Languages

Related East Germanic languages, spoken by tribes who emigrated from southern Scandinavia about the same time as the Goths and even earlier, include those of the Vandals (who eventually established themselves in North Africa) and of the Burgundians (who set up a kingdom in Gaul); both languages became extinct in the 6th and 7th centuries and almost nothing is known of them except personal and place names.

Influence of Gothic on Other Languages

The Visigoths engaged in missionary work, spreading their Arian version of Christianity (which held that Christ, though divine, was not equal with the Father) among other East Germanic tribes, for whom Gothic was apparently a lingua franca. For example, the Burgundians, Vandals, and Ostrogoths were converted in the 4th and 5th centuries. It has been claimed that this missionary activity led ultimately to the appearance of a few specifically Gothic linguistic forms in the Bavarian and Alemannic dialects of Old High German (a West Germanic language), and even as far afield as Old English. The route and extent of this influence, however, and the question of whether it was direct or indirect, are subjects of scholarly debate.

Later History and Crimean Gothic

Some Goths, forced out by invading Huns in the 3rd century, migrated westward and founded kingdoms in modern Italy (Ostrogoths), France, and Spain (Visigoths), but their power was shattered everywhere by the beginning of the 8th century and Gothic became extinct in the west.

The language lived on longer in the east, surviving in present-day northern Bulgaria until the 9th century and in the Crimea until the 16th century, according to accounts by travelers. One of the last of these was by Ogier Ghiselin (or Ghislain) de Busbecq, the imperial ambassador to the Ottoman court at Constantinople in 1560–1562, who recorded 68 Crimean Gothic words as well as some phrases and numerals, with Latin translations. This collection is of rather limited value, for it was published without his permission and may contain misprints; more importantly, his two informants were dubious, one being a native Greek speaker. A great lack is that of any indication of morphological variation and syntax. Nevertheless, it appears possible that Crimean Gothic was a descendant of a somewhat different variety of Gothic from that of Wulfila.

Busbecq also notated a short ‘Gothic’ song (the *Cantilena*) but he gives no translation and it has been variously claimed to be not Gothic but Turkish, Swedish or Italian.

By the end of the 18th century C.E. Crimean Gothic had apparently died out.

Bibliography

- Balg G H (1891). *The first Germanic Bible*. New York: Westerman.
- Bennett WH (1960). *An introduction to the Gothic language*. New York: Modern Language Association of America.

- Braune W (1980). *Gotische Grammatik* (19. Auflage), bearbeitet von Ernst A. Ebbinghaus. Tübingen: Niemeyer.
- Ebbinghaus E A (1992). 'Some remarks on the life of Bishop Wulfila.' *General Linguistics* 19, 15–29.
- Friedrichsen G W S (1961). *Gothic studies*. Oxford: Blackwell.
- Green D H (1998). *Language and history in the early Germanic world*. Cambridge: Cambridge University Press.
- Heinemeyer W (1961). *Studien zur Geschichte der gotischen Urkundenschrift*. Köln: Bohlaus.
- Marchand J W (1973). *The sounds and phonemes of Wulfila's Gothic*. The Hague: Mouton.
- Russell J C (1994). *The Germanization of early mediaeval Christianity*. London: Oxford University Press.

- Stearns M (1978). *Crimean Gothic: analysis and etymology of the corpus*. Saratoga, CA: Anna Libri.
- Streitberg W (1920). *Gotisches Elementarbuch* (6th edn.). Heidelberg: Winter.
- Streitberg W (1960). *Die gotische Bibel* (3rd edn.). Heidelberg: Winter.
- Voyles J B (1981). *Gothic, German, and northwest Germanic*. Wiesbaden: Steiner.
- Wright J (1954). *Grammar of the Gothic language* [with selected texts, 2nd ed., with supplement by Sayce O L]. Oxford: Clarendon Press.

Greek, Ancient

A Teffeteller, Concordia University,
Montreal, Canada

© 2006 Elsevier Ltd. All rights reserved.

External History

Attested from the 14th century B.C., Greek has continued in an unbroken line of development down to the present day, the 'ancient' period coming to a close around 300 A.D. with the end of Hellenistic Greek. In the 1700 years from the Mycenaean period to the *koinē* and beyond, the language underwent significant changes in phonology, morphology, syntax, and lexicon. A member of the Indo-European family of languages, Greek has particular affinities with Indo-Iranian; its connections with Latin, once thought to be so close, in fact largely reflect cultural interaction, rather than subgrouping features.

We do not know when the language entered Greek lands, but it was in use on the mainland and on Crete in the second half of the second millennium B.C. During the first millennium, it was spoken, in one form or another, on the Greek mainland; the Aegean islands, including Crete, Cyprus, and Rhodes; in parts of Asia Minor; and in southern Italy and northern Africa. From earliest times, Greek existed as a collection of dialects, with Attic, the dialect of Athens, eventually dominating and serving as the foundation for the Hellenistic *koinē* and its further development into later stages of the language.

Dialects

Mycenaean

Mycenaean Greek was written in a syllabic script on clay tablets used for record keeping in the Bronze Age

centers on the Greek mainland and on the island of Crete. Deciphered only in 1952, these documents, although they contain no literature (which presumably at this time was still maintained exclusively in an oral tradition) and (owing to the nature of their contents) offer only limited evidence particularly of syntax and verb morphology, nevertheless provide us with an invaluable source of information regarding the development of Greek from Proto-Indo-European. Although the Linear B syllabary does not mark all contrasts (e.g., one series of signs represents both [l] and [r], and there is no differentiation of [k], [k^h], and [g]; of [p] and [p^h]; or of [t] and [t^h], although both voicing and aspiration were phonemic in the language), the script does provide a series of signs for the labiovelars (PIE **k^w*, **g^w*, **g^wh*), which by the time of alphabetic Greek in the 8th century B.C. had developed into the phonologically conditioned series of stops [p, b, p^h/t, d, t^h/k, g, k^h].

'Historical' Dialects

The dialectal status of Mycenaean is disputed. We know that it was not the only variety of Greek spoken in the second millennium because it shows innovations not shared by all of the later-attested dialects, such as assibilation of *-t* before *-i*; debate continues as to whether it should be grouped unqualifiedly with any of the so-called historical dialects attested in the first millennium (and, with the exception of Cypriot, written in an alphabetic script), although its affinities with Arcado-Cypriot are clear. The dialects are grouped as follows: Attic-Ionic (in Attica and its chief city Athens, the Ionic islands of the Aegean, and parts of Asia Minor), Aeolic (including Boiotian, Thessalian, and Lesbian), Doric (or West Greek, in the Peloponnese, the Doric islands of the Aegean,

and parts of Asia Minor), Northwest Greek (on the northern mainland), and Arcado-Cypriot (in Arcadia and Cyprus). The old view, supported by the testimony of the ancient Greeks themselves, that the dialects entered Greece in three successive waves – Attic-Ionic, Aeolic with Arcado-Cypriot, and Doric with Northwest Greek – has recently been challenged by a model that locates the dialectal differentiation in Greece itself during the course of the second millennium.

‘Literary’ Dialects

Certain dialects or conventionalized forms of them were associated with certain genres of literature. Thus, regardless of the native dialect of a given author, dactylic hexameter poetry (epic, etc.) was predominantly in Ionic with a heavy admixture of Aeolic, choral poetry was in Doric, the dialogue of Athenian tragedy was in Attic (the choral lyrics were in Doric), and so on.

Phonology

Vowels

The phonological system included 10–12 vowels, subject to dialect variation: the short vowels [a, e, o, i, u] (with [u] fronted to [y] in Attic) and the long vowels [a:, e:, o:, i:, u:] (with [u:] fronted to [y:] in Attic). In addition, some dialects, including Attic, distinguished long open *e* and *o* (ε: and ο:), in addition to [e:] and [o:]. The situation was further complicated by the orthography (e.g., the long close *e* [e:] that arose from compensatory lengthening fell together with the original long *e* in some dialects and was written η, the pronunciation of which varied from higher to lower position in different dialects). Other dialects, which distinguished [e:] and [ε:], wrote the latter as η, the former as ει, a sign that had earlier represented the diphthong [ej] but had subsequently been reduced to a monophthong, thus allowing for the use of the digraph to represent compensatorily lengthened [e:] as well. Similarly, [o:] was written ου in those dialects that distinguished [o:] and [ο:] (ω). Four diphthongs remained in frequent use: [aj, oj, aw, ew] (written αι, οι, αυ, ευ). The status of υι is uncertain. The ‘long’ diphthongs [a:j, e:j, o:j] (written αι, ηι, ωι or, with iota subscript, a Byzantine innovation, αι, ηι, ωι) generally lost the glide or merged with the short diphthongs.

The Indo-European vocalic resonants **r*, **l*, **m*, **n*, yielded vowel reflexes or vowel and consonant combinations (e.g., **n* > α/αν, **r* > αρ/ρα/ορ/πο).

A major dialectal feature of Ionic and, with exceptions, Attic is the raising of inherited [a:] to [ε:] (written η).

Consonants

There were nine stops, grouped in three points of articulation – labial, dental, velar – with each point of articulation having three types: voiceless, voiceless aspirate, and voiced: [p, p^h, b/t, t^h, d/k, k^h, g]. There were four resonants – the liquids [l, r] and the nasals [m, n] – and two fricatives: the dental [s] (with a voiced allophone [z] before voiced consonants) and the glottal [h], the so-called rough breathing (marked ‘) in word-initial position before some vowels, e.g., Ἑλλάς (*Hellas*), ἅγιος (*hagios*) ‘holy’, and before *r*, e.g., ῥόδον (*rhodon* ‘rose’) (where it may indicate not aspiration but a voiceless pronunciation). Initial [h] arose as a development from earlier [s], [j], etc. The velar nasal [ŋ] occurred as an allophone of [n] before velars and was usually written γ, e.g., ἄγγελος (*angelos*), cf. Eng. ‘angel’. In the liquids λ represented a dental lateral [l], and ρ in non-initial position was apparently a voiced rolled *r*, as in, e.g., Italian. The three letters ζ, ξ, and ψ represented consonant clusters. The sound represented by ζ seems to have varied in different dialects and at different periods from [dz] to [zd], eventually simplifying to [zz] or [z]; ξ represented [ks] and ψ [ps]. The voiced semivowel [w], widely attested in Mycenaean, was dropped early in Attic-Ionic (cf. Myc. *wa-na-ka* = Hom. and Att. *anaks* ἄναξ), but was retained by some dialects well into the historical period, where it was represented by the letter digamma (Ϝ).

The loss of digamma and of intervocalic *s*, etc., resulted in the vowel contractions characteristic of the later language. Many morphophonemic changes resulted in complex developments, particularly of consonant clusters. Constraints on word-final consonants left only final *-r*, *-s*, *-n*. The aspirated stops [p^h, t^h, k^h] developed by the later period into the corresponding fricatives [f, θ, χ].

Accent

Ancient Greek had a pitch accent: rising (acute ‘), falling (grave `), or rising-falling (circumflex ^), the latter restricted to long vowels and diphthongs. The accent was ‘free,’ subject to certain phonological and morphological constraints; it fell only on one of the last three syllables of a word and was recessive in verbs (a residue of an early stage in which the verb of a main clause was cliticized).

Morphology and Syntax

Ancient Greek has a very rich derivational and inflectional morphology. Use is made of prefixes, very rarely of infixes, and overwhelmingly of suffixes.

Derivational processes depend primarily on composition and suffixation. The verbal morphology is especially complex.

Morphosyntax of Nominals

Ancient Greek has three declensional classifications for nouns and adjectives: *o*-stems, $\bar{\alpha}$ -stems, and consonant stems. There are five cases (nominative, genitive, dative, accusative, vocative), three numbers (singular, dual, plural), and three genders (masculine, feminine, neuter). There are remnants of an instrumental case (e.g., the suffix *-phi* in Mycenaean and Homeric Greek) and the locative (e.g., *oikoi* 'at home'); otherwise, the functions of the locative, instrumental, and ablative cases reconstructed for Proto-Indo-European are taken over by the genitive and dative in Greek. The dual category, already unstable in Homer, was eventually largely eliminated. Grammatical gender was not necessarily determined by natural gender; although terms for males and females were as a rule masculine and feminine, respectively, this was not exclusively the case, and inanimate objects could be categorized indiscriminately as masculine, feminine, or neuter.

Morphosyntax of Verbs

The verbal morphology encodes many morphosyntactic categories. Finite forms are marked for person and number; nonfinite forms include infinitives, participles, and other verbal adjectives. In addition to person and number, finite verbs are marked for tense, aspect, voice, and mood. The complexity of the Greek verbal system arises from its retention of the already complex system of Proto-Indo-European (PIE), to which is added a number of new categories of tense/aspect and voice (e.g., the future, the pluperfect, further contrasts of active, middle, and passive, etc.), and innovations in the aspectual status of the aorist and perfect. Thus, the PIE aspectual opposition of present/aorist/perfect, indicating, respectively, imperfective, perfective, and stative aspect, was maintained to a great extent, especially in the non-indicative moods and the nonfinite forms. However, in the indicative 'tenses,' the opposition of present/aorist/perfect was combined with distinctions of time that eventually overshadowed the original distinctions of aspect so that the aorist could be used simply as a past tense and the perfect came to develop a resultative use.

The tenses are present, imperfect, aorist, perfect, pluperfect, future, and future perfect. The categories of voice are active, middle, and passive (the middle indicating close involvement of the subject in the

action, e.g., reflexive usage, etc.). The moods are indicative, subjunctive, optative, and imperative.

The Indo-European process of morpheme-internal vowel gradation (*Ablaut*), e.g., *e/o/Ø*, etc., was widely used in Greek. What began as a phonological process was morphologized already in the pre-PIE period and yielded many distinctions in both nouns and verbs in Greek, for example, the distribution seen in the following forms built on the root **pet* 'fly': *pétomai* 'fly,' *poté* 'flight,' and *pterón* 'wing'; within the verbal paradigm, examples are present *peíthō* 'I persuade,' perfect *pépoitha* 'I am persuaded.'

Syntactic Typology

Ancient Greek has a relatively free word order that resists simple classification in terms of SOV, SVO, and so on. More fruitful is a recent approach (Devine and Stephens, 2000) that sees the syntactic typology of Ancient Greek as changing from an earlier non-configurational type to configurationality, from a syntax of juxtaposition to a syntax of government and embedding. Particularly in its earlier stages (and in archaizing poetic traditions), Greek made pervasive use of discontinuous constituency, adjunct lexical arguments, null anaphora, and parataxis. Thus, the language of the Homeric poems, which were recorded in writing in the 8th or early 7th century B.C., but represent a prior oral tradition of a thousand years or more, is heavily paratactic, whereas 4th-century Attic prose is, by contrast, markedly hypotactic, with complex forms of subordination.

Specific features of interest in Greek syntax include the use of particles, the distribution of clitics, and the development of the article.

Bibliography

- Allen W S (1987). *Vox Graeca* (3rd edn.). Cambridge: Cambridge University Press.
- Buck C D (1955). *The Greek dialects* (2nd edn.). Chicago: University of Chicago Press.
- Chadwick J (1963). 'The prehistory of the Greek language.' In Edwards I E S, Gadd C J & Hammond N G L (eds.) *Cambridge ancient history* (3rd edn.). Vol II. Cambridge: Cambridge University Press.
- Chantraine P (1961). *Morphologie historique du grec* (2nd edn.). Paris: Klincksieck.
- Devine A M & Stephens L D (1994). *The prosody of Greek speech*. New York and Oxford: Oxford University Press.
- Devine A M & Stephens L D (2000). *Discontinuous syntax: hyperbaton in Greek*. New York and Oxford: Oxford University Press.
- Dover K J (1968). *Greek word order* (2nd edn.). Cambridge: Cambridge University Press.
- Lejeune M (1972). *Phonétique historique du mycénien et du grec ancien*. Paris: Klincksieck.

- Meillet A (1975). *Aperçu d'une histoire de la langue grecque* (8th edn.). Paris: Klincksieck.
- Palmer L R (1980). *The Greek language*. London: Faber & Faber.
- Rix H (1992). *Historische Grammatik des Griechischen* (2nd edn.). Darmstadt: Wissenschaftliche Buchgesellschaft.

- Sihler A L (1995). *New comparative grammar of Greek and Latin*. New York and Oxford: Oxford University Press.
- Ventris M & Chadwick J (1973). *Documents in Mycenaean Greek* (2nd edn. with reprint of 1st edn.). Cambridge: Cambridge University Press.
- West M L (1982). *Greek metre*. Oxford: Oxford University Press.

Greek, Modern

B D Joseph, Ohio State University,
Columbus, OH, USA

© 2006 Elsevier Ltd. All rights reserved.

General Overview

Although very much a living and vibrant language with speakers numbering in the millions around the world, Modern Greek actually began to develop thousands of years ago, when speakers of the ancient form of the language entered the Balkan peninsula some time in the early part of the second millennium B.C. These speakers moved quickly, according to most current accounts, into the southern part of the region – what is now northern and central Greece and the Peloponnesos – and into most of the neighboring islands of the Aegean Sea, including Crete as the most southerly point. This settlement area essentially defines the space where to this day the Greek language remains an enduring presence, though there has been spread into other areas, in some cases dating from ancient times.

Modern Greek is the official language of the Hellenic Republic (i.e., the Republic of Greece) where there are some 11 000 000 speakers, and also of the Republic of Cyprus, with some 600 000 speakers. In large part because of ancient colonization, Greek is found today in numerous communities and enclaves around the Mediterranean and Black Sea area, including Sicily, southern Italy, Alexandria (Egypt), and the region around the Crimean peninsula. Moreover, Greeks in modern times have migrated to many locations throughout Europe (but especially England), Australia (with a large concentration around Melbourne), and North America (particularly in New York, Chicago, Ohio, Florida, and Toronto), forming the modern-day 'Hellenic Diaspora.' Although Greek is mainly a second language in these diaspora communities, it is still robust and alive there, and these communities add perhaps as many

as 2 500 000 speakers to the overall total of speakers of Greek worldwide.

The language is generally referred to as 'Greek' in English, but the linguistic autonym for Greek speakers is based on an entirely different root. Greek speakers call their language *eliniká* (i.e., 'Hellenic') or *neoeliniká* (i.e., 'neo-Hellenic'). Occasionally, the designation *roméika* is used also; it is literally, 'Romaic', a use deriving from the affinities many (Orthodox Christian) Greeks have felt for the Eastern Roman (or Byzantine) Empire, centered in Constantinople after the 4th century A.D.

The modifier 'modern' is generally used in referring to the language in English, in much the same way that the Greeks themselves often use *neo-*, literally 'new', in their self-designation (*neoeliniká*, as above). Indeed, the unadorned label 'Greek' in English usually refers to the ancient language. This usage recognizes the fact that the language has a long and rich documented history, being attested as early as the 14th century B.C. (so-called Mycenaean Greek) and continuing through ancient times and the Byzantine era up to modern times.

The modern form of the language is significantly different from its ancient Greek predecessor with regard to pronunciation and general structural features, but at the same time, as perhaps with all languages, there is noticeable continuity as well. The changes that set Modern Greek apart from the ancient language (e.g., the falling together of some eight distinct vocalic nuclei to [i], the shift from a pitch accent to a stress accent, a greater degree of analyticity in nominal and verbal constructions for earlier synthetic ones, among others) can be seen in nascent form in the period of the Hellenistic Koine. By the 10th century A.D., the language in many respects had a quite modern look to it. Still, it is customary to date the modern period of Modern Greek to approximately the 17th century, recognizing that even in the so-called Medieval Greek period, some structural differences from contemporary Greek were to be found

(e.g., syntactically in the continued use of an infinitive, morphologically in the formation of the future tense, and phonologically in the expansion of a dental affricate and the elimination of a front rounded vowel), as well as numerous lexical differences.

Dialects of Modern Greek

Taken as a whole, Modern Greek exhibits great diversity across all its varieties, defined both geographically and socially. However, the considerable differences are largely masked by the dominance and ubiquity of the standard language, the variety that reflects the everyday usage of speakers in Athens and its environs, by far Greece's leading population center, with over 4 000 000 inhabitants, and the country's focal point for culture, economy, religion, and government.

Looking first at diversity from a geographic standpoint, the major modern regional dialects (following Newton, 1972a) that can be identified are Peloponnesian-Ionian Greek, traditionally viewed as the basis for the contemporary standard language; northern Greek, in a zone starting north of Attica (where Athens is located) and extending up to and beyond Greece's second largest city, Thessaloniki; Cretan, the dialect of the island of Crete; Old Athenian, the dialect of Athens before the 1821 War of Independence and, as a result of various resettlements, found elsewhere in Greece into the early 20th century; and southeastern Greek, including Greek of the Dodecanese islands, as well as Cypriot Greek. However, modern Cypriot shows significant differences on all levels (phonological, morphological, and syntactic) that invite its classification as a separate language.

Two other important geographic varieties include (1) Tsakonian, the rather divergent form of the language, a direct descendant of the ancient Doric dialect that is spoken still in the eastern Peloponnesos; and (2) the Pontic dialects, which were once spoken along the Black Sea coast (Crimea area and Asia Minor), but are now mostly found in various parts of Greece as a result of the 1923 population exchanges with Turkey. Both Tsakonian and Pontic diverge significantly enough from the rest of Greek to merit consideration now as separate languages (though they are still clearly Hellenic).

Sociolinguistic Setting and Other Diversity

Geography and regional dialects account for only part of the diversity present in the Greek-speaking world; an additional crucial facet is the diglossia

(in the sense of Ferguson, 1959) that Greek exhibits, as an outcome of centuries of cultural influence from the Classical Greek language and Classical Greece itself on modern speakers. Classical Greek and Classical Greece were treated as the prescriptive norms against which speakers of later stages of Greek generally judged themselves, resulting in a 'two-track system' for the language, with a consciously archaizing form that speakers and writers modeled on Classical Greek set against a vernacular innovative variety. With the founding of the new nation-state of Greece after the revolution of 1821, these two tracks developed into a significant register and stylistic difference between a high-style variety associated with official functions (those involving government, education, religion, and the like), known as *katharevousa* ('Puristic', literally 'the purifying' language), and the ordinary, day-to-day language of the people, known as *dimotiki* ('demotic', literally 'the popular' language). These two varieties vied for status as the primary form of the language; each had its advocates, for whom language attitudes tended to correlate with certain social attitudes and political positions, more conservative for advocates of *katharevousa* and more progressive for followers of *dimotiki*. The competition continued throughout most of the 20th century, with *katharevousa* generally being in the ascendancy for official use, but was resolved most recently by various governmental acts and actions in 1976 declaring *dimotiki* as the official language. Still, all throughout the various official and unofficial periods of diglossia, the usage that speakers exhibited has actually been mixed, showing borrowing between the two varieties, in particular with *katharevousa* forms incorporated into *dimotiki*. The present state of Standard Modern Greek is essentially *dimotiki*, but with significant borrowings from *katharevousa* involving grammar (morphology and syntax), pronunciation, and vocabulary.

Linguistic diversity for Modern Greek, therefore, involves the mixing of varieties of both a regional and stylistic/social nature and mutual interactions among them.

Structure of Modern Greek

Modern Greek's vowel system is fairly unremarkable, showing /i e a o u/ with no distinctive length or nasality. Consonants include /p t k f v θ ð s z x ɣ j r l m n t̪ d̪/; /b d g/, although deriving in some analyses from underlying nasal plus voiceless stop combinations, are probably best taken as distinctive elements in their own right. Still, the consonants are somewhat overstocked with fricatives, by most typological standards.

Modern Greek has a distinctive stress accent, restricted to occurring only on one of the last three syllables in a word; to some extent, accent placement is tied to particular morphological categories, but in general there is some degree of unpredictability as to which of the final three syllables is to be stressed.

In its morphology, Modern Greek is for the most part synthetic and fusional, with grammatical endings marking two numbers (singular and plural) and four cases (vocative, nominative, accusative, and genitive, which covers some 'dative'-like functions) in the noun. In the verb, there is a complex interplay of realizations for tense (present, past, future), mood (indicative, imperative, subjunctive), aspect (perfective and imperfective, with the so-called perfect tense perhaps forming a third distinction), person (speaker, hearer, and other; that is, first, second, and third), and number (singular and plural). Endings carry most of the marking functions, but some categories are realized by prefixal (or prefix-like) elements, especially the future tense and subjunctive mood. Weak object pronouns under some analyses are considered to be transitivity markers on the verb, thus possibly constituting a further inflectional category. Negation too might be considered to be realized via prefixal elements.

With regard to syntax, Modern Greek shows a significant degree of analyticity, even with its generally synthetic morphology. Sentential complementation is always via person- and number-marked finite clauses, as there is no infinitive proper in the language, and for some case functions, especially of the genitive case, prepositional phrases occur as alternatives. Word order is fairly free, responding more to pragmatic and discourse-related criteria, such as focus and topicalization, than to purely syntactic concerns. Dislocated elements are often cross-indexed, so to speak, on the verb through the use of agreeing weak pronouns (so-called object reduplication).

Much of what Modern Greek shows in the way of surface syntactic and morphological patterns that differs from Ancient Greek may be attributable to interactions between speakers of Greek and speakers of other neighboring languages in the Balkans during the medieval (i.e., pre-modern) period, though language-internal factors clearly played a major role too.

Bibliography

- Browning R (1983). *Medieval and Modern Greek*. Cambridge: Cambridge University Press.
 Dawkins R (1916). *Modern Greek in Asia Minor: a study of the dialects of Silli, Cappadocia and Pharasa, with*

- grammar, texts, translations and glossary*. Cambridge: Cambridge University Press.
 Drettas G (1997). *Aspects pontiques*. Paris: Association de Recherche Pluridisciplinaires.
 Eleftheriades O (1985). *A grammar of contemporary Modern Greek*. Palo Alto, CA: Pacific Books.
 Ferguson C (1959). 'Diglossia.' *Word* 15, 325–340.
 Holton D, Mackridge P & Philippaki-Warbuton I (1997). *Greek. A comprehensive grammar of the modern language*. London: Routledge.
 Holton D, Mackridge P & Philippaki-Warbuton I (2004). *Greek: an essential grammar of the modern language*. London: Routledge.
 Horrocks G (1997). *Greek. A history of the language and its speakers*. London: Longman.
 Householder F, Kazazis K & Koutsoudas A (1964). *Reference grammar of literary Dhimotiki*. Bloomington, IN: Indiana University/The Hague: Mouton.
 Jannaris A (1968). *An historical Greek grammar*. Hildesheim: Georg Olms.
 Joseph B (1990). *Morphology and universals in syntactic change: evidence from medieval and Modern Greek*. New York: Garland Publishers.
 Joseph B (1999). 'Selected titles on language and linguistics.' In Constantinidis S (ed.) *Greece in modern times (an annotated bibliography of works published in English in 22 academic disciplines during the twentieth century)*. Lanham, MD: Scarecrow Press. 441–474.
 Joseph B & Philippaki-Warbuton I (1986). *Modern Greek*. London: Croom Helm.
 Koutsoudas A (1962). *Verb morphology of Modern Greek: a descriptive analysis*. The Hague: Mouton.
 Mackridge P (1985). *The Modern Greek language*. Oxford: Oxford University Press.
 Mirambel A (1961). *Introduction au grec moderne* (3rd edn.). Paris: G. P. Maisonneuve-Larose.
 Mirambel A (1969). *Grammaire du grec moderne*. Paris: Presses Universitaires de France.
 Moleas W (1989). *The development of the Greek language*. Bristol: Bristol Classical Press.
 Newton B (1972a). *The generative interpretation of dialect*. Cambridge: Cambridge University Press.
 Newton B (1972b). *Cypriot Greek. Its phonology and inflections*. The Hague: Mouton.
 Pernot H (1934). *Introduction à l'étude du dialecte tsakonian*. Paris: Société d'Éditions "Les Belles Lettres".
 Rohlf G (1948). *Historische Grammatik der unteritalienischen Gräzität*. Munich: Verlag der Bayerische Akademie der Wissenschaften.
 Seaman P (1972). *Modern Greek and American English in contact*. The Hague: Mouton.
 Shipp G (1979). *Modern Greek evidence for the Ancient Greek vocabulary*. Sydney: Sydney University Press.
 Sotiropoulos D (1972). *Noun morphology of modern Demotic Greek. A descriptive analysis*. The Hague: Mouton.
 Swanson D (1960). *Modern Greek studies in the West: a critical bibliography of studies on Modern Greek*

linguistics, philology, and folklore in languages other than Greek. New York: New York Public Library.

Thumb A (1964). *A handbook of the Modern Greek language: grammar, texts, glossary*. Angus S (trans.). Chicago: Argonaut Press.

Tonnet H (1993). *Histoire du grec moderne. La formation d'une langue*. Paris: L'Asiathèque.

Warburton I (1970). *On the verb in Modern Greek*. Bloomington, IN: Indiana University.

Guaraní

W F H Adelaar, Leiden University, Leiden, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Guaraní (or *Avañe'ẽ*) is the name of a language spoken in parts of the South American lowlands, in particular in the basin of the Paraná and Paraguay rivers. It is one of the official languages of Paraguay, where it is the mother tongue of a majority of the population. The present number of speakers of Guaraní is estimated at 5 000 000. It survives as a minority language in neighboring areas of Argentina (Corrientes, Misiones) and Brazil (Mato Grosso do Sul, Paraná). In Bolivia, Guaraní (*guaraní boliviano*) is the name used for the closely related Chiriguano language. Originally an Amerindian language, Guaraní is now used by most layers of the Paraguayan population, regardless of their Indian or mixed background. They often speak a variety of Guaraní with a heavy Spanish admixture, called *jopará*. More conservative varieties of Guaraní are spoken by tribal groups such as the Chiripá (called Nhandéva in Brazil), the Mbyá and the Paĩ-Tavyterá (or Kaiová). The purity of the Guaraní literary language is watched over by writers and academic institutions.

The Guaraní language is part of a larger family, Tupi-Guaraní, which has a wide distribution in the South American lowlands. The latter also includes Tupinambá, the language that was used as a general language (*língua geral*) along the coasts of Brazil before the 19th century. Tupi-Guaraní, in its turn, is part of a larger stock named Tupí, which may be distantly related to the Cariban and Macro-Ge families. The original area of diffusion of both Tupi-Guaraní and its Tupí sister languages appears to be the Guaporé basin in southwestern Brazil (state of Rondônia). The expansion of the Tupi-Guaraní peoples may have occurred shortly before the European conquest. The strong position of Guaraní in Paraguay is related to the consequences of Jesuit missionary policy during the 17th and 18th centuries. The first authoritative grammar of Guaraní was written by a Jesuit, Antonio Ruiz de Montoya (1640).

Modern Paraguayan Guaraní differs from the language described by Montoya in that it has lost some of its original morphophonemic complexity. Guaraní is a mildly polysynthetic language with a loosely structured morphology. It uses both prefixes and suffixes, the former being more tightly bound to the root than the latter. Guaraní prefixes indicate person (of subject, object, and possessor), mood (subjunctive, imperative), reflexive, reciprocal, accompaniment, and causative (except for transitive stems). There are two portmanteau prefixes that indicate the combination of a first person actor with a second person patient, e.g., *po-hecha* [poheʃa] 'I see you (plural)' (*po-* '1st person actor with 2nd person plural patient'). Suffixes indicate case, diminutive, collective, causative (of transitive stems), future tense, mood (wish, intention, etc.), nominalization, subordination, and several other verbal categories. Some grammatical categories (aspect, number, most tenses) are indicated by elements that are best interpreted as separate words, as in *o-mba'apo hína* [õmbaʔa'po hĩnã] 'he is working' (*o-* '3rd person actor; *hína* 'progressive aspect'). Negation is usually indicated by a combination of a prefix and a suffix, as in *nd-o-hó-i* [ndo'hoj] 'he does not go' (*nd(a)-..i(ri)* 'negation'; *o-ho* 'he goes'). Noun incorporation (both of subject and object) is frequent, especially in the colloquial language. Most incorporated nouns refer to body parts; e.g., *che-akā-rasy* [ʃēākāra'si] 'I have a headache' (*che-* '1st person subject'; *akā* 'head'; *(-r)asy* 'hurt').

Like many other South American lowland languages, Guaraní features a so-called active/stative system. That is, transitive verbs are inherently active, but intransitive verbs can be classified either as active or as stative; compare active *a-guata* [aɣʷa'ta] 'I walk' (*a-* '1st person singular actor') and stative *che-mandu'a* [ʃēmāndu'a] 'I remember' (*che-* '1st person singular subject'). Adjectives (quality verbs) can be treated as a subclass of the stative verbs. Possessed nouns can also express the notion 'to have'; e.g., *che-róga* [ʃe'roɣa] 'my house' (*che-* '1st person singular possessor,' *(-r)óga* 'house'), but also 'I have a house.' This makes it difficult to distinguish between stative verbs, on one hand, and expressions of possession, on the other.

Many nouns and stative verbs are subject to an alternation of their initial consonant (the most frequent set is *t-/r-/h-*) depending on the construction in which the form occurs; e.g., *téra* 'name,' *che-réra* 'my name,' *héra* 'his/her name.' Most prefixed forms take *-r-*; forms with initial *r-* also indicate the core of a genitive construction, in which the modifier precedes the modified; e.g., *yvága rape* [i'vaɣa ra'pe] 'the road to heaven' (*yvaga* 'heaven,' *tape* 'road').

Grammatical relations, except subject and inanimate object, are indicated by case markers, postpositions, or combinations of both. Case markers tend to merge with pronouns or pronominal prefixes into special forms; for instance, *ha'e* [ha'e] 'he, she, it' with ablative *-gui(ve)* [ɣ^wi'(ve)] is realized as (*i*)*chugui* [(i)ʃuɣ^wi]; with comitative *-ndi(ve)* [ndi'(ve)] as *hendive* [hēndi've], etc.

Guaraní is widely known from the linguistic literature as an example of the existence of prosodic nasality or nasal harmony (a common feature in Amazonian languages). It has six oral and six nasal vowels. Prosodic nasality takes its source from a (stressed) nasal vowel or a nasal consonant and spreads leftward covering the root that contains the source as well as all its prefixes. Suffixes show a more independent behavior: they may adapt to the root or maintain their own nasality/orality structure. When

nasality is generated by a nasal consonant (/m/, /n/, /ŋ/, /ŋ^w/) the latter marks the beginning of an oral domain, hence it is realized as half-nasal half-oral [mb, nd, ŋg, ŋg^w], e.g., in *marangatu* /maraŋa'tu/ [MĀRĀŋa'tu] 'holy' (the nasal domain is indicated in small caps). Guaraní roots generally have final stress. When stress is penultimate, rightward nasal spread can occur as well. Some suffixes are affected by rightward spread, e.g., *kuña* + *-pe* → *kuña-me* /ku'ñā-me/ [kūñāmē] 'to the woman' (*-pe* 'dative').

Bibliography

- Dixon R M W & Aikhenvald A Y (eds.) (1999). *The Amazonian languages*. Cambridge: Cambridge University Press.
- Gregores E & Suárez J A (1967). *A description of colloquial Guaraní*. The Hague, Paris: Mouton.
- Guasch A (1956). *El idioma guaraní. Gramática y antología de prosa y verso* (3rd edn.). Asunción: Casa América, Moreno Hermanos.
- Meliá B (1992). *La lengua guaraní del Paraguay*. Madrid: MAPFRE.
- Montoya A Ruiz de (1640). *Arte y vocabulario de la lengua guaraní*. Madrid (1876: facsimile edition by Julius Platzmann, Leipzig.).
- Velásquez Castillo M (1996). *The grammar of possession. Inalienability, incorporation and possessor ascension in Guaraní*. Amsterdam, Philadelphia: Benjamins.

Gujarati

B K Waghmar, University of London,
London, UK

© 2006 Elsevier Ltd. All rights reserved.

Gujarātī belongs to the southwestern family of Modern/New Indo-Aryan, a subgroup of the Indo-Iranian branch of Indo-European languages. The official language of Gujarat state, it is spoken across South Asia in Maharashtra (especially Bombay), Rajasthan, Sind, lower Punjab, Madhya Pradesh, and in Karnataka and among the Parsi, Hindu, Muslim, and Jain diaspora in the Persian Gulf, East and South Africa, Britain, North America, and Australia. There were approximately 45 479 000 speakers reported in 1997 (*Indian Missions Abroad*).

History and Literature

Scholars historically distinguish Old Gujarātī (12th–15th centuries); Middle Gujarātī (15th–18th centuries); and Modern Gujarātī (18th century onward).

Its antecedents are traceable to a distinct Old Western Rajasthānī literary form, despite the attestation of Jain Prākṛit treatises and studies by Middle Indian grammarians of *Nāgara Apabhraṃśa*, a literary *Apabhraṃśa* of Gujarat. The 12th-century *Bharateś-varabāhubalirāsa* (1185) is the earliest work written in Gujarātī. Prose and verse compilations written from the 13th century onward exist and include the seasonal poem *Vasantavilāsa* and the 14th-century commentary, the *Ṣaḍāvaśyakabālabodhavr̥tti*. Narasiṃha Mehtā's (c. 1414–1480) devotional ballads marked a new era in poetry, acquiring pride of place in its literary annals. The Gujarātī daily, *Mumbai Samācār* (established in 1822), is one of the oldest newspapers in Asia. Bombay Parsis were pioneers in Gujarātī and Urdū theater from the 1850s.

Dialects

Gujarātī spoken along the Baroda-Ahmedabad corridor is regarded as the standard/prestige dialect. (Whether the register of Nāgarī Brāhmaṇs carries

‘RP’ status remains debatable). Other dialects are Surati (southern Gujarat), Carotarī (Charotari; central Gujarat), Kāthiāwārī (Saurashtra), and Pāṭānī (northern Gujarat). Pakistani Gujarātī is probably a Pāṭānī subdialect, and code switching is waning as the younger generation shifts to Urdū and provincial languages. Muslim speakers there and elsewhere obviously adopt Perso-Arabic lexicons – its largest word stock after Sanskrit – especially in religious-cultural discourse. Parsi Gujarātī, an ethnolect of the subcontinent’s Zoroastrians is, however, readily intelligible. East African Gujarātī now contains Swahili loanwords. Kacchī (Kachchi) is semantically intermediate between Gujarātī and Sindhī and is also influenced by Mārwarī.

Grammar

Phonetically, Gujarātī is unique for murmured vowels developed from final /h/ and two open vowels, /ɛ/ and /ɔ/. An absence of contrast exists between short and long /i/ and /u/ vowels. Variable or invariable substantives and adjectives, as well as pronominals, have three genders (including the neuter) and two numbers; they inflect for direct and oblique forms, the latter with post-positions and clitics. Verbal forms have temporal, modal, and aspectual contrasts. Combinations of verbal nouns and adjectives with auxiliaries produce an elaborate variety of obligational and desiderative forms, and the vocabulary is rich in passive, causative, and double causative verbs (Cardona, 1965). Vector/compound verbs, a common New Indo-Aryan feature, are employed in restricted contexts with specific semantics.

Orthography

A manuscript dated 1592 (Mistry, 1996) attests that an alphasyllabic script derived from a Devanāgarī variant has been employed for writing Gujarātī and Kacchī since the 16th century. A cursive style replaced the standard Sanskrit script used in prose and verse when printing began during the 1830s. Independent and conjunct forms are expressed by 45 symbols: 8 vowels, 34 consonants, *anusvāra*, *visarga*, and a velar nasal grapheme. Written from left to right, Gujarātī is conspicuous for its absence of head strokes and varying phonemic modifications. As in other Brāhmī-derived scripts, the post-consonantal /a/ is evidently assumed in a consonant lacking diacritics.

Devanāgarī-derived numerals were adopted with modified shapes for the digits 3, 5, 6, and 9.

Bibliography

- Acharya S (1966). *Kacchī śabdāvalī* [Kutchi vocabulary]. Ahmedabad: Gujarāt Vidyāpīth.
- Acharya S (1985). *Halari dialect*. Ahmedabad: Gujarāt Vidyāpīth.
- Bhandari A (1990). *Gujarātī vākyaracanā* [Gujarati syntax]. Ahmedabad: University Grantha Nirmāṇa Board.
- Bhayani H (1988). *Gujarātī bhāṣhanum aitiḥāsika vyākaraṇa* [A historical grammar of Gujarati]. Gandhinagar: Gujarāt sāhitya akādamī.
- Cardona G (1965). *A Gujarati reference grammar*. Philadelphia: University of Pennsylvania Press.
- Cardona G & Suthar B (2003). ‘Gujarati.’ In Cardona G & Jain D (eds.) *The Indo-Aryan languages*. London: Routledge. 659–697.
- Dave J (1995). *Colloquial Gujarati: a complete language course*. London: Routledge.
- Dave R (1977). *Studies in Gujarati phonology and phonetics*. Ph.D. diss., Cornell University.
- Dave T (1935). *A study of the Gujarati language in the 16th century*. London: Royal Asiatic Society.
- Deshpande P & B (1988). *Universal English-Gujarati dictionary*. Bombay: Oxford University Press.
- Dwyer R (1995). *Gujarati: a complete course for beginners*. London: Hodder Headline PLC.
- Gajendragadkar S (1974). *Parsi-Gujarati: a descriptive analysis*. Bombay: University of Bombay.
- Joshi U, Raval A & Shukul Y (eds.) (1973–1981). *Gujarātī sāhityano itihāsa* [History of Gujarati literature] (4 vols). Ahmedabad: Gujarātī sāhitya pariṣad.
- Lambert H (1971). *Gujarati language course*. Cambridge: Cambridge University Press.
- Mamaeva V (1990). *Гуджарати-русский словарь: Gujarātī-Ruśīyan śabdkoś* [Gujarati-Russian Dictionary]. Moscow: Russkii yazyk.
- Master A (1940–42). ‘Some English loan-words in Gujarati.’ *Bulletin of the School of Oriental and African Studies* 10, 25–30.
- Mistry P (1969). *Gujarati verbal constructions*. Ph.D. diss., University of California at Los Angeles.
- Mistry P (1996). ‘Gujarati writing.’ In Daniels P & Bright W (eds.) *The world’s writing system*. New York & Oxford: Oxford University Press. 391–394.
- Naik C (1972). *Gujarātīmām pracalita phārasī vyutpatti kośa: arabī tatha turkī muḷanā śabdo sabita* [Persian etymons prevalent in the Gujarati language together with words of Arabic and Turkish origins]. Ahmedabad: Gujarat University.
- Učida N (1990). *A Saurashtra-English dictionary*. Wiesbaden: Harrasowitz.

Gullah

T Hopkins, Florida International University, Miami, FL, USA

© 2006 Elsevier Ltd. All rights reserved.

Gullah is the name of the language spoken by the former African slaves and their descendants (also called Gullahs) living along the southeastern U.S. Atlantic coastline from the northern tip of North Carolina to Jacksonville, Florida, especially in the Sea Islands off the coast of South Carolina and Georgia and the bordering mainland areas (e.g., the low country). The Gullah language and its speakers are also respectively known as Geechee and Geechee(s), a local in-group term, and because of its negative connotation it is not used by outsiders. Other contemporary terms used to refer to the Gullah language and its speakers have included Sea Island Creole and Sea Islanders, but these terms have largely been restricted to the academic literature. Another more recent term that has sometimes been used to identify the Gullahs is Native Islanders, employed largely on Hilton Head Island to distinguish the native Gullah Sea Islanders from the influx of outsiders that began settling on the sea islands in the 1950s. However, this term has not received widespread acceptance among the Gullah population and Gullah still seems to be the preferred ethnic classification for the language and its speakers. Gullah has historical and linguistic links with Afro-Seminole, an offshoot of Gullah spoken by black Seminoles and their descendants, who were forced out of northern Florida to the west in the 1830s, and are now living in central Oklahoma, western Texas, and northern Mexico. The exact number of Gullah speakers today is uncertain because no census has attempted to distinguish the Gullahs from other African Americans living in the region.

History

The Gullahs are descendants of Africans brought from the Caribbean and directly from West Africa to provide the slave labor force for a plantation agricultural system, supported by English planters who began settling in the low country region in the late 1600s. The low country's rich fertile soil was ideal for the prosperous production of such plantation crops as rice, indigo, and the long-stapled Sea Island cotton for which the Sea Islands became known. The success of the plantation system was owed not only to the slave labor force but also to the agricultural and technical skills of the Africans, particularly in rice cultivation. Following the U.S. Civil War in 1865, the English planters abandoned their plantations, leaving the

freed Africans and their descendants to eke out a living of their own. They engaged in small-scale farming, hunting, and fishing, and the relative isolation of the Sea Islands allowed them to preserve many of the traditions and customs of their African ancestors. Today, fish netting, basket weaving, woodcarvings, and quilting are some of the customs that characterize the rich cultural lifestyle of the Gullah people.

The etymology of the term Gullah has not yet been determined. Gullah may have derived from Ngola, the name of an African tribe in the Hamba basin of Angola, a West African region from which many of the Africans were transported, or it may have originated from Gola, the name of an African tribe and language near the Liberia-Sierra Leone border in West Africa. The etymology of Geechee is also unknown. Geechee may have derived from the Ogeechee River plantation in Georgia, the word *gidzi* in Mende to mean a country called Kissy (Liberia), or it may have come from the name of a tribe and language in Liberia. Equally obscure is how or when Gullah became the local ethnic classification for the enslaved Africans and their descendants living in the low country and their language. Historical citations show its first appearance in the literature in 1822, when the Charleston City Council records made reference to Gullah Jack and his company of Gullah or Angola Negroes. This indicates that its early use was associated with a particular group of Africans. However, the Gullah people and their language are far more diverse as both their historical development and language reveal.

Language Development

Gullah is the only English-based creole in the United States, although its origins are still speculative. Three hypotheses have been proposed of its development. One hypothesis suggests that Gullah developed from the Caribbean English spoken by the Africans and their descendants on plantations in the Caribbean and was brought with the African slaves when transported to the low country and subsequently learned and linguistically influenced by the Africans who were later imported into the region. Another hypothesis is that Gullah originated in a pidgin that developed on the West African coast and was brought with Africans imported to the low country. A third hypothesis suggests that Gullah developed on the plantations in the low country out of contact between English-speaking settlers and the various West African languages of the African slaves. Whatever the source, the retention of Africanisms in all components of the Gullah grammar is a common thread, and it is largely

these features, commonly associated with an African substratum, that is often used to distinguish Gullah from its English lexifier source.

Linguistic Characteristics

Gullah differs from General American English most obviously in its phonology. Some of these features include the use of the voiced bilabial fricative [β] or the voiced labiodental frictionless continuant [v] for the voiced labiodental fricative [v] in English, e.g., [βeri] cf. 'very,' [səvænə] cf. 'Savannah'; the use of the voiced palatal nasal [ɲ] in, e.g., [ɲustə] cf. 'used to'; and the absence of interdental fricatives in, e.g., [tɪn] cf. 'thin' and [di] cf. 'the.' The prenasalized and labiovelar stops [kp/gb] were reported in early Gullah phonology but these sounds were largely used in the Africanisms, and although Africanisms are still present in the Gullah lexicon, they do not appear to be in widespread use among the Gullah today.

The Gullah lexicon is largely English. From his research conducted in the late 1930s, Lorenzo Turner was the first linguist to document over 4000 Africanisms in the Gullah lexicon, many of them used as basket names (e.g., Gullah nicknames). Today you can still hear in normal everyday conversations such African retentions as *buckra* 'white man,' *tita* 'elder sister,' *dada* 'mother or elder sister,' *nyam* 'eat/meat,' *sa* 'quickly,' *benne* 'sesame,' *una* 'you,' and *da* the verb 'to be.' Other Gullah Africanisms such as *cooter* 'turtle,' *tote* 'to carry,' *okra* 'plant food,' *gumbo* 'stew,' and *goober* 'peanut' are widely used in mainstream American English. In addition to direct loans, the Gullahs also employ a large number of calques, e.g., *day clean* 'dawn,' *day broad clean* 'full daylight,' *first dark* 'sunset,' *night shut-in* 'midnight,' *sweet mouth* 'flattery,' *bad mouth* 'to denigrate,' *i foot broke* 'to become pregnant,' *big eye* 'greedy,' and *hard head* 'dumb.'

Its most interesting features morphosyntactically include the pronominal form *unə* 'you' (2nd person sg/pl); *nə* 'and' (conj); *də* 'the verb be'; *bɪn* 'was'; *sɛ/fə* 'that' (complementizer); *dɛm* to refer to '& company' when postposed to proper animate nouns as in *Suzi dem* 'Susie and others' and 'those' as a demonstrative determiner to common nouns as in *dem boy* 'those boys.' Like its English lexifier creoles, Gullah also employs several preverbal auxiliaries within its tense, aspect and modal verbal paradigm. The primary auxiliaries are *də*, *ə*, *bɪn*, *don*, *dəz*, and *go*, which can occur alone or in combination with each other to express a wide range of tense, modal, and aspect

meanings in past/nonpast contexts. These same forms function as lexical verbs. As an auxiliary [*də*] in *I də go* generally indicate durative and habitual meanings cf. 'I am/was going' and 'I generally go,' respectively. Habitual meaning is also expressed with the auxiliary [*dʌz*]; the auxiliary [*də*] can also express iterative and perfective meanings. Its phonological variant [*ə*] indicates durative, habitual, iterative, and perfective meanings. The auxiliary [*don*] expresses perfective meaning in *I don go* cf. 'I have gone'; the auxiliary [*bɪn*] in, e.g., *I bɪn go* indicates anterior meaning cf. 'I went/had gone'; the auxiliary [*go*], as well as auxiliary [*ə*], expresses future meaning as in *I go go* or *I ə go* cf. 'I will/would go.' Gullah is not homogeneous, so the use of these grammatical forms and meanings may vary depending on the region (e.g., the sea island or mainland area the Gullah speaker originated) and sociocultural factors (e.g., age, education, socioeconomic status, etc.).

In the mid-1950s, developers launched the Sea Islands on an explosive growth in development and population called progress, which led to an unprecedented assault on the relatively stable Gullah people and their way of life. The Sea Islands, which had until then been isolated, were now connected by causeways and bridges to the mainland (only two islands, Daufuskie Island, South Carolina, and Sapelo Island, Georgia, still remain unattached to the mainland but have not escaped development). The possible extinction of the Gullah population was threatened and local organizations, such as the Penn Center, the Gullah Coalition, and Gullah Festival, were inaugurated and are now contributing to the preservation of the Gullah language and cultural identity.

Bibliography

- Hancock I (1980). *The Texas Seminoles and their language*. Austin: University of Texas at Austin.
- Hopkins T (1992). Issues in the study of Afro-Creoles. Ph.D diss., Indiana University. Ann Arbor: University Microfilms International.
- Hopkins T (1994). Variation in the use of the auxiliary verb *dc* in contemporary Gullah. In Montgomery M (ed.) *The Crucible of Carolina*. Athens: The University of Georgia Press.
- Joyner C (1984). *Down by the Riverside: a South Carolina slave community*. Urbana: University of Illinois Press.
- Smith R (1969). 'Gullah.' Bulletin of the University of South Carolina. Columbia: University of South Carolina. A Reprint of Bulletin No. 190 of the University of South Carolina, 1969.
- Turner L D (2002). *Africanisms in the Gullah dialect*. Columbia: University of South Carolina Press.

Gur Languages

J Bendor-Samuel, Summer Institute of Linguistics,
High Wycombe, UK

© 2006 Elsevier Ltd. All rights reserved.

There are some 85 Gur languages within the Volta-Congo branch of the Niger-Congo family. These languages are used in an area that stretches from southeast Mali across the northern Ivory Coast, a large part of Burkina Faso, all of northern Ghana, northern Togo, northern Benin, and into northwest Nigeria. They predominate within the parallels 6°W and 2°E and 8°N and 14°N, the savannah lands north of the forest belt where the Kwa languages are spoken. The name 'voltaïque' is used in French.

The Speakers

The number of speakers of Gur languages is probably in the region of 12 to 15 million people. There is a wide range in the size of the Gur language groups; some are large, such as the Moore in eastern Burkina Faso numbering some four million; others are much smaller with only a few thousand speakers.

Gur Studies

The earliest record of Gur languages is found in S.W. Koelle's *Polyglotta Africana*, published in 1854, which includes word lists from 10 Gur languages. There was, however, little further mention of Gur languages until the 20th century. Westermann (1927) recognized the validity of Gur as a subfamily. Greenberg (1963) and Bendor-Samuel (1971) confirmed this classification and added considerable detail to earlier work.

The first in-depth study of any Gur language was the extensive work done by Manessy (1975, 1978) and Prost in the 1970s. The Summer Institute of Linguistics began research in Gur languages in northern Ghana in the 1960s; their efforts have produced several studies in recent years (see Naden, 1989).

Languages and Their Classification

Most of the Gur languages belong to two main groups – Central Gur and Senufo – but there are several languages that do not fall into either group. Furthermore, although there is little doubt about the grouping together of the languages in Central Gur and similarly of the languages in Senufo, this bringing together of these two major groups of languages

into a single subfamily, Gur, is being increasingly questioned. However, because no other relationship to any other language group is any closer, Senufo remains within Gur for the time being.

Central Gur languages are found in northern Ghana, eastern Burkina Faso, and northern Togo and Benin and break down into two main clusters, Northern and Southern. In the Northern cluster, the Oti-Volta group predominates, with 27 out of 29 languages. These 27 languages can be subdivided into six unequal clusters with the largest, known as Western, having 13 languages, whereas the next largest, Gurma, comprises six languages. The Oti-Volta group includes a number of major languages, such as Moore, Dagaari, Dagbani, Frafra, Mampruli, Kusal (Kusaal), and Konkomba. It is noteworthy that the linguistic relationship among languages in the Oti-Volta group is considerably closer than that among languages in the Grusi group.

In the Southern cluster, the Grusi group comprises 20 of the 29 languages. The Grusi group is more scattered, with the seven eastern languages of the group being completely separated from the four northern and nine western languages by a substantial block of Oti-Volta languages.

Central Gur includes another dozen languages outside the Oti-Volta and Grusi groupings.

The Senufo group comprises 20 languages, of which 10 are related more closely to each other and form a subgroup, Senari. The Senufo languages are found on the western side of the Gur group in the northern Ivory Coast and southwestern Burkina Faso.

A further nine Gur languages do not belong to any of the groupings so far set up within Gur. Detailed studies of many more Gur languages are needed so that the relationships of the languages to each other can be established more clearly.

Structural Features

Phonetics and Phonology

Although there is a great deal of diversity among the Gur languages, a subgroup, such as Oti-Volta, includes languages with many features in common.

Consonant sets often include voiced and voiceless plosives and fricatives and nasals at five points of articulation: labial, alveolar, palatal, velar, and labiovelar, plus l, y, and w. Phonetic [r] often occurs as a non-initial allophone of /d/.

The geographically contiguous northern and eastern Grusi languages display a vowel harmony system. A systematic contrast in vowel length is common throughout the Gur languages, as is the syllabic nasal.

All Gur languages are marked by contrastive tones, the domain of which is usually the syllable, but may be the word or the morpheme; these tones usually carry grammatical rather than lexical implications. There is no one tonal system; both terrace systems with two tones and downstep and systems with up to four contrastive pitch levels are found.

Both CV and CVC roots occur, as well as C(V) suffixes. In many languages, transitional vowels are inserted to avoid consonant clusters.

Grammar and Syntax

Most Gur languages have singular and plural class suffixes, and in some languages, pronouns and other NP elements concord with the head noun. The contrast between imperfective and neutral forms is very common and usually occurs by means of verbal extensions, though a tone change or vowel lengthening is also found. Other verbal categories are usually shown by particles, auxiliaries, or occasionally tone.

Aspect rather than tense is marked, though 'past' may be contrasted with 'nonpast' or 'future' with 'nonfuture,' and there are often time-depth particles

(e.g., 'one day away' may be 'tomorrow' with future or 'yesterday' with nonfuture).

SVO is the predominant word order in Gur languages, though Senufo languages show SOV word order.

Bibliography

- Bendor-Samuel J (1971). 'Niger-Congo, Gur.' In Sebeok T A (ed.) *Current trends in linguistics 7: Linguistics in Sub-Saharan Africa*. The Hague: Mouton. 141–178.
- Greenberg J H (1963). *The languages of Africa*. The Hague: Mouton & Co.
- Manessy G (1975). *Les langues oti-volta*. Paris: Centre Nationale de la Recherche Scientifique.
- Manessy G (1978). 'Les langues voltaïques.' In Barreteau D (ed.) *Inventaire des études linguistiques sur les pays d'Afrique noire d'expression française et sur Madagascar*. Paris: Conseil Internationale de la Langue Française. 71–83.
- Naden T (1989). 'Gur.' In Bendor-Samuel J (ed.) *The Niger-Congo languages*. Lanham and London University Press of America, Inc. 141–168.
- Westermann D (1927). *Die westlichen Sudansprachen und ihre Beziehungen zum Bantu*. Hamburg: Reimer.

Guugu Yimithirr

J B Haviland, Reed College, Portland, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

The Language and Its Speakers

Guugu Yimithirr (Gu:guyimidjir) (hereafter GY) is the language originally spoken in the area between the Annan and Jeannie Rivers on the coast of north-east Queensland and inland. Most of its modern speakers now cluster around Cooktown, at the mouth of the Endeavour River. The language name combines *guugu* 'word' or 'language' with the comitative *yimi-thirr* 'this way' – thus, 'this kind of language' or 'speaking this way.' (Contrast the name of its southern sister Kuku-Yalanji, which has *yala* for GY *yii* 'this, thus.'). GY contributed Australia's most widespread loanword to the languages of the world, via the word *gangurru* [IPA g'aŋuru] 'large grey wallaroo,' which was recorded as 'kangaroo' by members of Captain Cook's crew during their stay on the shores of the Endeavour in 1770. Speakers of the language were spared further contact with Europeans for about 100 years, when gold was discovered inland at the Palmer River in the mid-1870s. The consequent gold rush and invasion of the territory by settlers decimated

local Aboriginal populations, and within 10 years the few surviving speakers of GY lived in scattered hunter-gatherer bands pushed off their lands or on the fringes of Cooktown and other smaller towns. A Lutheran mission established on barren land at Cape Bedford in 1885 became a refuge for the remaining GY speakers, and GY was the lingua franca of the community as legislation relocated Aboriginal children from a wide area – including many who spoke different languages – at the mission school. The earliest written information about GY derives from the Cape Bedford missionaries, systematized by W. E. Roth, the first Northern Protector of Aborigines. The 20th century saw a severe GY diaspora, as speakers of the language who had migrated to Cooktown and Cape Bedford were forcibly relocated to southern Queensland during World War II, and returned to their homeland in only a fraction of their already reduced numbers in the 1950s. Nowadays, though most GY speakers still live around Cooktown, others are scattered through other Queensland Aboriginal communities, and as far away as Melbourne and New Zealand. Despite repeated predictions, starting in the 1920s, that GY was on the verge of extinction, the language remains a central feature of life at the Hopevale community north of

Cooktown, and there are currently more than 1000 speakers of the language, which has undergone rapid change over the past half-century.

Genetic Relations

GY is fairly typical of northern Paman languages, and is very closely related to Kuku Yalanji, still a relatively healthy language spoken in the rainforest to the south, with which it has perhaps 40%–50% lexical overlap. GY has a fairly typical Australian phonemic inventory. Although most stems are disyllabic and consonant initial, GY is closely related to the now extinct ‘initial dropping’ languages originally spoken to the north at Barrow Point and Flinders Island. Early accounts describe two varieties of GY, one *thalun-thirr* or ‘coastal’ and the other *waguurr-ga* ‘inland,’ though most diagnostic coastal lexical items have given way in modern speech to forms judged to be from the inland dialect. The language is now frequently written in a practical orthography which distinguishes two laminal series of stops and nasals (*th* and *nh* for laminodentals [d̪] and [ɳ], and *j* and *ny* for laminopalatals [ɟ] and [ɲ], respectively), although earlier orthographies still survive, including one developed by missionaries which omits laminal (and other) contrasts entirely. Rhotics contrast between an alveolar trilled *rr* and a retroflex *r* [ɭ]. There is a typical three-way vowel contrast between *i*, *u*, and *a*, and vowel length is phonemic. There is further a complex interaction between vowel length, canonical stress and syllable patterns, and suffixing morphology. (Long vowels are doubled in the practical orthography; in this article a colon marks a ‘lengthening’ suffix, and a dollar sign a shortening one – see below.)

Basic Facts

GY is wholly suffixing, with free personal pronouns (which follow a nominative/accusative case-marking pattern – contrasting with ergative/absolutive morphology on other nominal elements) and no bound pronominal forms. Pronouns, normally only referring to animate entities, distinguish three persons, three numbers (singular, dual, and plural), and, for some speakers, also an inclusive/exclusive distinction in non-singular first person. The case system is elaborate, with a variety of ‘local’ (locative/allative, ablative [= causative], superessive, abessive, adessive [= goal]) and peripheral syntactic cases (dative, purposive, instrumental) in addition to those marking the core syntactic relations of subject and object. This case system is elaborated still further with the morphologically hypertrophied cardinal direction roots, detailed below.

Genitive constructions allow dual case marking, where a possessed nominal functions in some further case-marked role, as in the following examples (which also show the ergative/absolutive patterning of nouns and the nominative/accusative pattern of pronouns).

- (1) nyulu yarrga thadaara biibaa-:ga-mi.
 3*S.NOM* boy.*ABS* go.*REDUP* father-GEN-
 NONP LOC/ALL
 ‘the boy is going to his father’s place.’
- (2) jirraayn-da nyulu ngaliinh bama
 man-ERG 3*S.NOM* 2*DU.ACC* *man.ABS*
 ngarrbal daama-y
 strange.ABS *spear-PAST*
 galga thawuuny-ga-mun
 spear *friend-GEN-INST*
 ‘the man speared me and a stranger with a friend’s
 spear.’

Verbs fall into three main conjugations, with several minor additional patterns. Morphologically productive verbal categories are again numerous, including past and nonpast tenses, repetitive and continuous aspects, and such moods as contrafactual, desiderative, cautionary, precautionary or 'lest' forms, and a morphologically incorporated negative (in addition to a periphrastic negative construction). A multifunctional verbal suffix *:-thi* marks a syntactically complex reflexive/reciprocal construction, as outlined below.

Word order is completely free, and there is no discernable favored or unmarked order for clausal or, indeed, phrasal constituents. Each word of a discontinuous nominal constituent, including determiners and adjectives, can be marked individually for case. Animate noun phrases frequently appear with both a full nominal head as well as the appropriate case form of an apparently pleonastic pronoun, so that the nominative/accusative case-marking pattern of pronouns frequently coexists within the same clause with the ergative/absolutive morphology on other nominal constituents.

Shortening vs. Lengthening Suffixes and the Rhythmic Canons

GY shows an intriguing interaction between suffixation, vowel length, and stress. Virtually all monosyllabic lexical words have stressed long vowels, and long vowels are usually also stressed. The language seems to favor a syncopated syllabic pattern, with alternate syllables short and long, unstressed and stressed. Thus, for example, the progressive aspect is formed by a complex partial reduplication of disyllabic verb stems to produce trisyllabic stems, whose middle syllable is a lengthened (and stressed) version

of the original second syllable. Thus, *balgal* ‘he’ll make it’ vs. *balgaalgal* ‘he is making it.’ In addition, suffixes differ according to how they affect stress and length in the stem to which they attach. Most suffixes engender length (when it is not already present) on all disyllabic stems ending in a nonnasal consonant. Thus, *gambul* ‘belly’ + *-hi* ‘LOC’ → *gambuulbi* ‘in the belly.’ Other ‘lengthening suffixes’ additionally engender length even on disyllabic vowel-final stems. Thus, *yugu* ‘fire’ + *:-ngu* ‘PURPOSIVE’ → *yuguungu* ‘for the fire.’ Still other ‘shortening’ suffixes shorten the long vowel in the second syllable of a disyllabic stem. Thus, *buurraay* ‘water’ + *:-\$a* ‘PURP’ → *buurraya* ‘for water.’

The Morphology and Semantics of Cardinal Directions

Rather than base locative expressions on body-relative or egocentrically anchored perspectives, GY, like many other Australian languages, uses locational descriptors which insistently incorporate cardinal directions. Four lexical roots, *gungga-*, *jiba-*, *naga-*, and *guwa-*, correspond roughly to the English directions north, south, east, and west, respectively, except that the GY terms denote compass quadrants rather than idealized points. In virtually all circumstances, GY speakers keep track of cardinal orientation and incorporate the appropriate directional terms into descriptions of both distant places and immediate locations. In answer to a question like ‘Where are you going?’ one will answer, for example, *nagaar bayan-bi* ‘east to the house.’ To tell someone to move a bit ‘that way’ one must add the correct direction: *yarrba guwa-manaayi* ‘move a bit that way to the west’ (literally, ‘thus west-be’).

Whereas ordinary nominal expressions have just a single LOCATIVE/ALLATIVE form, and another ABLATIVE form, the directional roots have more elaborated morphological possibilities. The LOC/ALL forms, for example with the root *naga* ‘east,’ number three:

- (3) LOC/ALL forms
naga (0-form, ‘east from a point’)
naga-ar (R-form, ‘to a point east’)
naga-alu (L-form, ‘east, over some point or obstacle’)

Though all denote, in this case, motion in an easterly direction from some origo, each incorporates a different perspective or set of locational presuppositions. The least marked 0-form concentrates on the starting point of the trajectory, or emphasizes setting out toward the east. The second R-form focuses on the end point of the trajectory, also in the east, or emphasizes arrival. The third L-form is the most highly marked, presupposing some known or inferable

location to the east through or beyond which the current trajectory is conceived to pass. Similar elaboration extends to ablative and other locational case morphology, providing delicate resources for GY directional precision.

The Syntax and Semantics of the ‘Reflexive’ Suffix *:-thi*

GY transitive verbs ordinarily require animate subject NPs, whose referents are in an agent thematic role, conceived of as consciously and voluntarily controlling an action performed on some normally distinct object – the theme or patient argument.

- (4) *nyundu* *minha* *wagi* *naaybu-unh*
 2s.NOM *meat.ABS* *cut.PAST* *knife-INST*
 ‘you cut the meat with a knife.’

GY has a productive ‘reflexive’ construction, with the lengthening suffix *:-thi*, which encodes a basic variant of this situation, when agent and patient arguments are coreferential:

- (5) *ngayu(-ugu)* *wagi-ithi* *naaybu-unh*
 1s.NOM(-EMPH) *cut-REFL* *knife-INST*
 ‘I cut myself with a knife.’

Interestingly, GY uses the same verbal inflection, with varying case forms on the accompanying arguments, to encode other sorts of situation which depart from the canonical transitive situation characterized above. Thus, for example, in a situation appropriate to what in other languages might be encoded by a passive construction – for example, when there is no agent, or when the agent only accidentally acts, or when the organization of the discursive context promotes the object of the action to a position of prominence – GY uses the same *:-thi* suffix on the verb.

- (6) *nganhi* *wagi-ithi* *naaybu-unh*
 1s.ACC *cut-REFL* *knife-INST*
 ‘I got cut on the knife (by accident).’

Similarly, in a kind of generalized action in which no specific agent can be singled out, GY also has recourse to *:-thi*. A typical example might be

- (7) *nyulu* *gunda-athi*
 3s.NOM *hit-REFL*
 ‘he had a fight/was in a fight.’

There is also a small group of GY verbs which occur **only** in ‘reflexive’ form with *:-thi*, mostly denoting actions typically performed without conscious outside agency (‘come to an end,’ ‘explode,’ ‘finish,’ among others).

Language Situation and Sociolinguistic Features

Because of the particular history of its speech community, and the rapidly shifting conditions under which it is learned and spoken, GY is a language in a dramatic state of flux and variation. The mix of 'tribal' origins of its modern-day speakers, and the range of circumstances in which it serves as a medium of interaction, have produced different levels or registers in which GY and different varieties of Aboriginal and standard English combine.

Traditionally in this part of the Cape York peninsula, Aboriginal people were polyglots, often practicing linguistic exogamy and able to communicate as they traversed dialect and language areas. Even within single dialects, other socially significant linguistic varieties, such as the so-called mother-in-law or brother-in-law languages, provided linguistically marked ways of displaying deference to certain classificatory kinsmen, or of marking intimacy with others. Historically, as speakers of different Aboriginal languages (as well as creolized varieties of other contact languages) either congregated or were forcibly brought together at the Cape Bedford mission, GY became the native language of many people who still had ancestral ties to other 'tribal' languages. As knowledge faded of these other languages, so too did the specialized subvarieties of GY disappear, since they were systematically linked to social practices and processes of transmission which were radically altered by the sometimes violent upheavals in Aboriginal society.

In modern Hopevale, and around Cooktown, where the great majority of current GY speakers live, the language is still widely used, although it has a diglossic functional relationship with English. In the somewhat anarchic conditions of language acquisition in this fragmented speech community, the language is also undergoing probably accelerated simplification, as paradigms once fraught with irregularity are allowed to conform to more productive morphosyntactic patterns. Moreover, different generations in the community, with different kinds of schooling and a variety of personal backgrounds and competence in Australian English, mix English and GY freely in a typified and self-identifying variety of Hopevale

English which combines GY pronouns and individual lexical items with a largely English syntax.

Bibliography

- Dixon R M W (1980). *The languages of Australia*. Cambridge: Cambridge University Press.
- Haviland J B (1974). 'A last look at Cook's Guugu Yimithirr wordlist.' *Oceania* 44(3), 216–232.
- Haviland J B (1979a). 'Guugu Yimithirr.' In Dixon R M W & Blake B (eds.) *Handbook of Australian languages*. Canberra: Australian National University Press. 27–182.
- Haviland J B (1979b). 'Guugu Yimithirr brother-in-law language.' *Language in Society* 8(3), 365–393.
- Haviland J B (1979c). 'How to talk to your brother-in-law in Guugu Yimithirr.' In Shopen T A (ed.) *Languages and their speakers*. Cambridge, MA: Winthrop. 161–240.
- Haviland J B (1982). 'Kin and country at Wakooka outstation: an exercise in rich interpretation.' *International Journal of the Sociology of Language* 36, 53–70.
- Haviland J B (1985). 'The evolution of a speech community: Guugu Yimithirr at Hopevale.' *Aboriginal History* 9(1–2), 170–204.
- Haviland J B (1991). '"That was the last time I seen them, and no more": voices through time in Australian Aboriginal autobiography.' *American Ethnologist* 18(2), 331–361.
- Haviland J B (1993). 'Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures.' *Journal of Linguistic Anthropology* 3(1), 3–45.
- Haviland J B (1996). 'Projections, transpositions, and relativity.' In Gumperz J J & Levinson S C (eds.) *Rethinking linguistic relativity*. Cambridge: Cambridge University Press. 271–323.
- Haviland J B (1997). 'Owners vs. bubu gujin: land rights and getting the language right in Guugu Yimithirr country.' *Journal of Linguistic Anthropology* 6(2), 145–160.
- Haviland J B (1998a). 'Guugu Yimithirr cardinal directions.' *Ethos* 26(1), 1–23.
- Haviland J B (1998b). *Old Man Fog and the last Aborigines of Barrow Point*. Washington, DC: Smithsonian Institution Press.
- Haviland J B (2000). 'Pointing, gesture spaces, and mental maps.' In McNeill D (ed.) *Language and gesture: window into thought and action*. Cambridge: Cambridge University Press. 13–46.
- Haviland L K & Haviland J B (1980). 'How much food will there be in heaven? Lutherans and Aborigines around Cooktown before 1900.' *Aboriginal History* 4(2), 118–149.

H

Hausa

P J Jaggar, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Hausa is a Chadic (Afroasiatic) language spoken by an estimated 30 million or more first-language speakers (more than any other sub-Saharan language), mainly in northern Nigeria and southern Niger. It is also spoken by diaspora communities of traders, Muslim scholars, and immigrants in (mainly) urban areas of west Africa (e.g., Ghana, Cameroon), and also in the Blue Nile Province of the Sudan. Hausa is the most important and widespread west African language and continues to expand as a transnational lingua franca.

Hausa is used extensively in commercial, governmental, and educational spheres, and in the mass media. There are a number of Hausa-language newspapers, and book publishing, television, and video production are active. Many radio stations, both African and international, broadcast in (mainly Standard Kano) Hausa, including the BBC World Service, Voice of America, Radio Deutsche Welle, and China Radio International. A number of universities in Nigeria and Niger offer undergraduate and postgraduate degree courses in Hausa, and there are specialists in Hausa language and/or literature involved in comparable programs at universities in Europe, the United States, Japan, China, and South Korea. It has the best dictionaries (Bargery, 1934; Abraham, 1962; R. M. Newman, 1990) and reference grammars (Caron, 1991; Wolff, 1993; P. Newman, 2000; Jaggar, 2001) of any African language.

Substantial borrowing from neighboring African languages, such as Kanuri (central), the Mande group, Tuareg (Tamahaq, Tahaggart), Yoruba and Fula(ni) a single language (with considerable dialectal variation), has enriched the Hausa lexicon. Most loanwords come from Arabic, English, and French, however, with Arabic loans encompassing such semantic fields as religion (Islam was introduced to the area more than 500 years ago), education, government, law, commerce, war, and horsemanship.

Over the past 100 years, an ever-growing number of loanwords from English (Nigeria) and French (Niger) have been incorporated, typically denoting material objects and technology, education, and governmental and military positions.

Hausa has been written for more than 200 years in Arabic script ('àjàmi; see below for transcription), a system prevalent in Koranic schools and still used by many (mainly non-western-educated) Hausas for religious and literary purposes. However, 'àjàmi has been gradually supplanted by a modified Latin script/ alphabet called bookòo (probably < English 'book'), which does not, however, mark contrastive tone or vowel length. Hausa dialects vary in phonology, lexicon, and grammatical morphemes, and can be broadly grouped into Eastern Hausa (e.g., Kano = Standard Hausa, the variety described here), Western Hausa (e.g., Sokoto), and dialects in Niger (e.g., Aderanci).

Phonology

Hausa (Standard/Kano) has 32 consonant and 12 vowel phonemes (10 short/long monophthongal pairs plus two diphthongs) (See Table 1).

Table 1 Hausa: consonants

| Consonants: | | | | | | | |
|-------------|---|----|----|----|---|----|----|
| vl | f | fy | t | c | k | kw | ky |
| vd | b | | d | j | g | gw | gy |
| gl | ɓ | | d' | 'y | ƙ | ƙw | ƙy |
| vl | | | s | sh | | | |
| vd | | | z | | | | |
| gl | | | ts | | | | |
| | m | | n | | | | |
| | | | l | | | | |
| | | | r | | | | |
| | | | ř | | | | |
| | | | | y | w | h | |

c [tʃ] and j [dʒ] = palato-alveolar affricates, sh = palatal fricative [ʃ]; = glottal stop (orthographic ' '); ɓ, d' = laryngealized (implosives), 'y = glottalized palatal glide, ƙ [k'] and ts [s'] = glottalized ejectives; r [ɾ] = retroflex flap, ř = alveolar tap/roll.

Vowels (i = short; ii = long): i, ii, e, ee, a, aa, o, oo, u, uu.

Diphthongs: ai, au.

Hausa has two level tones, (H)igh (unmarked a[a]), and (L)ow (indicated with a grave accent on the first vowel à[a]), e.g., *kiifii* ‘fish’, *màcè* ‘woman’, *roogòo* ‘cassava’, *sàuka* ‘get down’. There is also a contour (H̄L) Falling tone (indicated with a circumflex), which occurs on heavy syllables, e.g., *râi* ‘life’, *kwântaa* ‘lie down’. There are three syllable types: CV, CVV (long vowel or diphthong), and CVC (long vowels are forced to conform and are reduced in extra-heavy *CVVC syllables). When followed by a front vowel, the coronal stops *t*, *d* and fricatives *s*, *z* palatalize to *c*, *j*, and *sh*, *j* respectively, e.g., *mèe sukà sàataa?* ‘what did they steal?’, *sun sàaci jàkaataa* ‘they stole my bag’ (< *sàati), *sun sàacee tà* ‘they stole it’ (< *sàatee), *taasàa* ‘metal bowl’ → pl. *taasooshii* (< *taasoosii).

Morphology

Within the pronominal system, the basic cut is between the personal and nonpersonal sets. There are eight sets of personal pronouns, each of which comprises eight forms: five singular (masculine/feminine gender is distinguished in the second and third person singular), and three plural. Personal pronoun paradigms express various syntactic functions, e.g., independent, (in)direct object, possessive, and reflexive. Nonpersonal pronouns (marked for gender and number) include demonstratives, interrogatives, and indefinites.

The categories of tense-aspect/mood (TAM) and subject agreement (person, gender, number) are marked on a lexically independent preverbal ‘person-aspect complex’ (INFL), which includes an additional (ninth) plural impersonal form (usually with arbitrary human reference and equivalent to a null subject). The TAM morphemes can be either segmentable, e.g., (Habitual) *ya-kàn zoo* ‘he regularly comes’, or fusional, involving changes in tone and/or vowel length, e.g., (Perfective) *yaa zoo* ‘he came’ and (Subjunctive) *yà zoo* ‘he should come’, and the subject-agreement elements are morphologically related to the personal pronouns. There are 16 distinct affirmative and negative inflectional paradigms, with a major Perfective:Imperfective aspectual dichotomy.

Hausa verbs are categorized into seven basic and derived ‘Grades’ (Parsons, 1960), defined in terms of their morphology (a templatic tone pattern and termination) and argument structure, and resembling the *binyanim* verbal conjugations in distantly related Semitic languages. Of the basic Grades (0–3), Grade 2 verbs (two-syllable) are canonically LH tone and are exclusively transitive, with two core (nonoblique) arguments, e.g., (Hausa is SVO) *Muusaa yaa ’àuri yaarinyā* ‘Musa married the girl’, *Muusaa yaa*

’àuree tà ‘Musa married her’. These examples also demonstrate how the final vowel of a Hausa verb can undergo changes in quantity and/or quality conditioned by the word class of the following constituent, e.g., direct object noun or pronoun. The derived Grades (4–7) add their own valency and semantics to the core meaning of the base form, e.g., basic Grade 2 verb *sàyaa* ‘buy’ → derived Grade 4 *sayèe* ‘buy up’, Grade 5 *sayař (dà)* ‘sell’, Grade 6 *sayoo* ‘buy and bring’, and Grade 7 *sàyu* ‘be completely bought up’. In the Imperfective, verbs are often replaced by participial-like verbal nouns, e.g., *sunāa daawôo-waa* ‘they are coming back’ (< Grade 6 *daawoo* ‘come back’ + ‘wà- with floating L tone), *mèe kakèe sàyee?* ‘what are you buying?’ (< Grade 2 *sàyaa* ‘buy’ [Grade 2 verbal nouns are nonpredictable]).

Nouns are masculine or feminine gender in the singular (an inherited Afroasiatic feature), and gender is overtly marked. Most native Hausa nouns are vowel-final. Feminine nouns typically end in -aa, e.g., *yaarinyāa* ‘girl’, *hùulaa* ‘cap’, or suffix-(I)YAA or -(U)WAA, e.g., *beebiyaa* ‘deaf mute (FEM)’, *gurgūwaa* ‘a cripple (MASC)’. Masculine nouns display a full range of final vowels (including -aa) and consonants, e.g., *raamii* ‘hole’, *beenee* ‘upper story’, *dilaa* ‘jackal’, *yaaròo* ‘boy’, *’àbù* ‘thing’, *mùtùm* ‘man’, *kaamùs* ‘dictionary’ (< Arabic).

Hausa noun (and adjectival) plurals are known for their complexity and involve suffixation, vowel insertion, and tonal melodies. They can be distilled into about 10 core classes, with the plural formation partially predictable from the canonical shape (e.g., tones, syllable structure) and, sometimes gender, of the singular. A disyllabic feminine singular with HL tone and final -aa, for example, will typically select a plural with the -ooCii suffix (where C = copy of final consonant of singular stem) and an all H tonal template, e.g. *jiikāa* ‘grandchild’ → pl. *jiikookii*. A disyllabic singular with a heavy CVV initial syllable and HH tones, on the other hand, is likely to pluralize by suffixing -àayee with HLH tones on the output, e.g., *giwāa* ‘elephant’ → pl. *giwāayee*.

Reduplication is pervasive, including (a) copying of a single consonant, e.g., past participial adjectives add a (MASC) suffix -aCCee, where CC = geminate copy of the stem-final consonant of the source verb, i.e., *cikakkee* ‘full, complete’ < *cikāa* ‘to fill’; (b) prefixal reduplication of the initial CVC- syllable of a sensory noun to form an intensive sensory adjective, e.g., *zàzzaafaa* ‘very hot’ (< *zaaf-zaaf-aa*) < *zaafii*; ‘heat’ (with gemination/assimilation of the coda C/f/); and (c) full reduplication (tones and segmentals), e.g., *coocì-coocì* ‘churches’ < *coocì* ‘church’, *yāu-yāu* ‘this very day’ < *yāu* ‘today’.

Syntax

The basic word order is S-V-IO-DO (goal/recipient arguments precede theme arguments), e.g., (ditransitive clause).

- (1) *daalibín yaa kaawoo*
student.the 3MASC.SG.PERF bring
wà maalàminsà 'aikii
to teacher.his work
 'the student brought the work to his teacher'

In *wh*-questions, relative clauses, and focus constructions, displaced constituents are moved to clause-initial position and special 'focus' marking is triggered on the INFL (Hausa is discourse-configurational), e.g., (FOC-PERF = Focus Perfective).

- (2) *wàa kukà gani*
who 2PL.FOC-PERF see
'à kàasuwa? (*wh*-question)
at market
 'whom did you see at the market?'

- (3) *yaarònkà nee mukà*
boy.your COP 1PL.FOC-PERF
gani (*ex situ* focus answer)
see
 'it was your boy we saw'

- (4) *yaaròn dà mukà*
boy.the REL 1PL.FOC-PERF
gani (relativization)
see
 'the boy that we saw'

Hausa is a pro-drop language, in that sentences can occur without overt NP subjects, e.g., [\emptyset]_{subj} [*sun*]_{infl} *tàfi gidaa* 'they have gone home', and it also licenses discourse-linked null (direct) objects, e.g., 'ii, *naa sàyaa* \emptyset 'yes, I bought (it)' (where \emptyset = null anaphor).

Negation in verbal sentences normally requires a double negative construction with the discontinuous morphemes *bà(a)* . . . *ba*, where the initial *bà(a)* occurs left-adjacent to INFL (following any overt subject), and the second is clause-final (though adverbs can occupy end position), as in (5) and (6).

- (5) *màataataa bà tà daawoo*
wife.my NEG 3FEM.SG.PERF return
bà (Negative Perfective)
 NEG
 'my wife has not returned'

- (6) *bàa zaa sù yāɗda ba* (Negative Future)
 NEG FUT 3PL agree NEG
 'they will not agree'

The Negative Imperfective uses only an initial High tone marker, as in (7).

- (7) *baa saa zuwāa*
 NEG 3PL.IMPERF coming
 'they are not coming'

Within the NP, grammatical gender and number trigger agreement on: (a) prehead elements, including adjectival modifiers ([8], also posthead), indefinite determiners (8), interrogative determiners, and universal quantifiers (also posthead); as in (8) and (b) posthead elements, including definite determiners, demonstratives (also prehead), relatives, numerals, genitive phrases, e.g., possessive NP, as in (9).

- (8) *wani kâramin*
 INDEF.MASC.SG small.MASC.SG.of.MASC.SG
yaaròo
boy.MASC.SG
 'a certain small boy'

- (9) *rùga-ĩ Mammàn*
gown.FEM-of.FEM Mammàn
 'Mammàn's gown'

Like many African languages, Hausa has a lexically autonomous class of highly expressive, phonosemantic words known as 'ideophones,' which normally function and distribute (sometimes collocationally) like adverbials. Hausa ideophones have their own distinct phonological and phonotactic properties (e.g., final obstruents, anomalous tones), and are semantically marked. As sound-symbolic elements, they typically denote the intensity or manner of an action, event, or state, e.g., (distinctive sounds or visually distinctive actions or features) *kwàngàràm* 'with a clang', *shaɗaf* 'soaking wet', *řasha-řasha* 'all sprawled out', and can also function as adjectives, e.g., *daɓaɓa* 'pronounced (facial markings)', and *fat* 'pure (white)'.

Constructions lacking a canonical verbal element include clauses containing an Imperfective INFL and nonverbal predicate, as in (10) and (11).

- (10) *Hàliimà tanàa dà*
Halima 3FEM.SG.IMPERF with
mootaa (possessive)
car
 'Halima has a car'

- (11) *sunàa masallaacii* (locative)
 3PL.IMPERF mosque
 'they are at the mosque'

Nonverbal clauses without any form of INFL include those in (12), (13), and (14).

- (12) *àkwai/baabù ruwaa nân* (existential)
 EXIST/NEG.EXIST water here
 'there is/is not water here'

- (13) gàa littaafinkà (presentational)
 PRES book.your
 'here/there is your book'
- (14) shii maalàmii nèe (equational/copular)
 3MASC.SG teacher COP.MASC.SG
 'he is a teacher'

Bibliography

- Abraham R C (1962). *Dictionary of the Hausa language*. London: University of London Press.
- Bargery G P (1934). *A Hausa-English dictionary and English-Hausa vocabulary*. London: Oxford University Press.
- Caron B (1991). *Le haoussa de l'Ader*. Sprache und Oralität in Afrika, 10. Berlin: Dietrich Reimer.

- Jaggar P J (2001). *Hausa*. London *Oriental and African Language Library*, vol. 7. Amsterdam & Philadelphia: John Benjamins.
- Jaggar P J & Wolff H E (2002). *Chadic and Hausa linguistics: the selected papers of Paul Newman with commentaries*. Cologne: Rüdiger Köppe.
- Newman P (2000). *The Hausa language: an encyclopedic reference grammar*. New Haven: Yale University Press.
- Newman R M (1990). *An English-Hausa dictionary*. New Haven: Yale University Press.
- Parsons F W (1960). 'The verbal system in Hausa.' *Afrika und Übersee* 44, 1–36.
- Wolff H E (1993). *Referenzgrammatik des Hausa*. *Hamburger Beiträge zur Afrikanistik*, vol. 2. Münster & Hamburg: Lit.

Hawaiian

P Geraghty, University of the South Pacific, Suva, Fiji

© 2006 Elsevier Ltd. All rights reserved.

Hawaiian belongs to the Eastern Polynesian branch of the Oceanic subgroup of the Austronesian language family, its nearest relatives being Tahitian and Marquesan.

Until the early 19th century it was spoken by the entire population of the Hawaiian islands, and it remained the main language for most of that century. However, the decline of the indigenous population, massive immigration, educational policies, and annexation by the United States (in 1898) have taken their toll, and Hawaiian was replaced by Hawai'i Creole English in the early 20th century as the main language of the native Hawaiian population. It is now estimated to have less than 1000 first-language speakers, mostly elderly individuals and the residents of the small island of Ni'ihau, out of a total population of over one million. Given the popularity of Hawaiian music among tourists, it could be said that Hawaiian is a language more sung than spoken.

Over the past 30 years grassroots moves to revive and revitalize the language, particularly through Hawaiian medium education from preschool (*pūnana leo*) to tertiary level, have had considerable success, and the number of speakers is increasing. Critics, however, point out that the style of Hawaiian spoken by this new generation who learned it in school is rather different from that of native speakers. It is now estimated that there

are 3000 fluent speakers of Hawaiian. There is very little Hawaiian-language radio and TV programming, and no newspaper.

Hawaiian had no traditional written form, and was first recorded by Captain Cook and his companions in 1778. A Roman-based alphabet was devised by English-speaking missionaries in 1829, and has remained in use relatively unchanged. In the 19th century, Hawaiians developed a high level of literacy through Hawaiian medium schools, and there is a large body of Hawaiian language literature from that period.

The phoneme inventory of Hawaiian consists of eight consonants (h, k, l, m, n, p, w, and glottal stop) and 10 vowels (a, e, i, o, u, ā, ē, ī, ō, ū). There are no consonant clusters and syllables are open. In writing, vowel length and glottal stop have often not been marked systematically. Most modern writers and publishers use a macron to indicate a long vowel and an inverted apostrophe (or apostrophe) to indicate the glottal stop.

There is very little morphophonemics, and most grammatical functions are performed by affixation or the use of pre- and postposed particles. Pronouns distinguish four persons (including first person inclusive and exclusive) and three numbers (singular, dual, and plural). There are two categories of possession, depending largely on whether or not the possessor has control over the fact of possession. In noun phrases, the order is head + attribute. The basic word order is VSO:

ua inu au i ka wai wela
 asp. drink I obj. the water hot
 'I drank the hot water'

Bibliography

Elbert S H & Pukui M K (1979). *Hawaiian grammar*. Honolulu: University Press of Hawaii.

Pukui M K & Elbert S H (1986). *Hawaiian dictionary*. Revised and enlarged edn. Honolulu: University of Hawai'i Press.

Schütz A J (1994). *Voices of Eden*. Honolulu: University of Hawai'i Press.

Hawaiian Creole English

S Romaine, Oxford University, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Hawai'i Creole English (locally called 'Pidgin') is the first language of the majority of locally born children and the first language of somewhat less than half the state of Hawai'i's population of just over a million. Varieties of pidgin and creole English in Hawai'i arose from contact between Hawaiians, Europeans (primarily English speakers, who contributed most of the vocabulary to the emerging pidgin), and the various immigrant groups (e.g., Chinese, Japanese, Portuguese, and Filipinos) brought to Hawai'i to work as indentured laborers on plantations from the 1850s onward.

Increasing contact between creole speakers and speakers of mainland varieties of English after World War II and the political incorporation of the islands into the United States as the 50th state in 1959 have blurred the boundaries between Standard English and the creole, and created a continuum of varieties. Although adjacent varieties of the continuum are mutually intelligible, the two extreme endpoints may often not be. This example illustrates the variation.

ai go give om da book fo yu most creole-like
ai gon give om da book fo yu
ai going give om da book fo yu
aim gonna give om da book fo yu
ail give om/him/her/them the book fo yu least creole-like
I will give him/her/them the book for you most standard English-like

The most decreolized (i.e., most English-like) varieties are found on the island of O'ahu, where three-fourths of the state's population is located, along with the capital, Honolulu, and the main U.S. military base, Pearl Harbor. The outer islands of Kaua'i and Hawai'i are the least decreolized.

Although most of the vocabulary of Hawai'i Creole English is derived from English, as many as a

thousand Hawaiian words may have been in use at one time during the plantation era; several hundred were probably in fairly common use colloquially. Although this number is now fewer, partly due to the decline in knowledge of Hawaiian, many still persist in local English and many more in the Hawai'i Creole English of older speakers. Younger people tend not to know the meanings of some words, such as *pio*, 'turn off, extinguish', and *kumu*, 'girlfriend, sweetheart', but virtually every resident locally would know *ono* 'tasty'. Many of these Hawaiian words are being replaced by English ones, although in some cases the English variants are still different from mainland U.S. English, e.g., *grinds* 'food' (cf. *kaukau* 'to eat/food').

Distinctive grammatical features include the lack of the copula *be* (*you da boss*, 'you are the boss'), use of *get* in both possessive and existential constructions (*get one wahine she get one daughter*, 'There is a woman who has a daughter'), use of *stay* for locatives and progressives (*Leilani stay inside da classroom*, 'Leilani is inside the classroom', *Charlene stay working*, 'Charlene is working'), use of *wen* or *had* as a simple past tense marker (*Joe wen/had talk to da coach*, 'Joe talked to the coach'), use of *pau* (Hawaiian 'done, finished') as a completive marker (*Call me when you pau*, 'Call me when you are finished'), preverbal negation (*Stan no mo rice*, 'Stan doesn't have any more rice'), and use of *for* as a complementizer (*Darrell like know how fo play basketball*, 'Darrell wants to know how to play basketball'). Word order is generally subject-verb-object, as in Standard English, apart from topic/comment structures such as *big, da house*, 'The house is big', in which the comment appears first.

There is considerable variation in pronunciation among local residents of different ethnic and social class backgrounds. Generally speaking, however, the phonology of Hawai'i Creole English has a smaller inventory of distinctive sounds than many mainland varieties of English. The [r] after a vowel in words

such as *shark* is usually absent in the most creole-like speech varieties (i.e., ‘park’ [pak]), and many of the diphthongs (double vowels) found in mainland varieties of English in words such as *coat* [kout] and *day* [dei] are single vowels in the creole (i.e., [kot], [de]). Hawai’i Creole English often has full vowels where mainland varieties use reduced ones, e.g., [tudei], ‘today’, vs. mainland [tɪde], and ‘mountain’ [mauntɪn] vs. mainland [mauntən]. English interdental fricatives in words such as *they* and *think* tend to become stops in the creole (i.e., [de], [tink]). The stops in consonant clusters such as [tr] in words such as *try* are affricated (i.e., [tʃrai]), and initial [s] in clusters such as [str] in words such as *stress* sounds more like [ʃ]. There are also some stress and intonational differences, such as the use of falling pitch for yes/no questions, which also do not show the subject/auxiliary inversion typical of standard English, e.g., *you like go Honolulu?* ‘Do you want to go to Honolulu?’ The falling intonation pattern has been carried over from Hawaiian into creole. Rising pitch together with a final question particle are used as a confirmation check in utterances such as *no mo job fo you, aeh?* ‘There isn’t a job for you, right?’

Despite the lack of written norms, standardization, or any official recognition, there are nevertheless some writers who have attempted to use the creole as a medium for poetry, short stories, and drama by adapting English spelling to represent some of the distinctive characteristics of speech varieties in Hawai’i. Each writer has worked out his or her own *ad hoc* spelling system. This burst of literary creativity can be seen partly as a manifestation of opposition to colonialism and as an affirmation of distinctive local identity in which the use of creole plays a key role.

Bibliography

- Reinecke J E (1969). *Language and dialect in Hawaii: a sociolinguistic history to 1935*. Honolulu: University of Hawaii Press.
- Romaine S (1994). ‘Hawai’i Creole English as a literary language.’ *Language in Society* 23, 527–554.
- Sato C J (1991). ‘Language attitudes and sociolinguistic variation in Hawaii.’ In Cheshire J (ed.) *English around the world*. Cambridge: Cambridge University Press. 647–663.
- Siegel J & Sakoda K (2003). *Pidgin grammar: an introduction to the creole language of Hawai’i*. Honolulu: Bess Press.

Hebrew, Biblical and Jewish

J F Elwolde, Valladolid, Spain

© 2006 Elsevier Ltd. All rights reserved.

Hebrew is the language of the people who, from the 11th century B.C., were dominant politically and culturally in and around what is the State of Israel. Members of this people stopped living in most of its ancient lands in any significant numbers following the Roman suppression of the Bar-Kochba revolt in 135 A.D. Indeed, for some 500 years previously Jews (from Judah/Judea, in the south of Israel) had settled in large numbers throughout the Hellenistic and Persian world. Even though they retained a high level of ethnic exclusivity, Jews became identified, and tended to identify themselves, as being more a ‘religious’ than a ‘national’ group, and this has been of enormous significance for the survival of Hebrew over three millennia.

A Holy Language

From the early centuries A.D., Hebrew is frequently referred to as *‘lēshōn ha-qōdesh* ‘the sacred language’ in contrast to other languages spoken by

Jews – particularly Greek, Aramaic, and, much later, Yiddish. In addition, the Hebrew of the Bible was sometimes regarded (by certain medieval writers, for example) as ‘purer’ than later varieties. It is possible that this view is attested as early as the second century B.C., when the Qumran community derided their religious adversaries as speaking an ‘uncircumcized tongue’ (Rabbinic Hebrew, perhaps). Jewish traditions claim that Hebrew was the language spoken at the creation, that the letters of the Hebrew alphabet were active in the creation, and that it is the only language understood by the angels and efficacious in prayer.

Biblical Hebrew

The main corpus of ancient Hebrew is the Bible (minus its Aramaic portions), consisting of 304 901 graphic word-tokens. The consonantal text of this corpus had achieved roughly its current state by the 2nd century B.C., after several revisions of its different parts. A point to be stressed is that Biblical Hebrew, as it has been handed down, comprises a literary, rather than an oral or ‘colloquial,’ corpus. The Bible is not a straightforward record of spoken

utterances from biblical times, but is better viewed as a collection of written compositions, including literary versions of conversations, orally transmitted stories, etc.

The language of those parts of the Bible composed after the Jews' return from exile in Babylonia (538 B.C.) is sometimes called 'Late Biblical Hebrew,' and it differs from the 'Classical' literary language of before the exile (586 B.C.). Partly it represents a consciously 'archaizing' imitation of the earlier language and partly, like Rabbinic Hebrew, a distinct, naturally developed, stage in the history of Hebrew.

When the Bible is viewed as a single work, albeit a composite of different works, it is clear that it has been written, or at least edited, from a religious perspective and with religious motives. Even so, it contains relatively little material that was composed in order to express feelings of a specifically religious character. Far more representative are historical or epic narratives, historical fiction, social polemic ('prophecy'), and detailed regulations about law, the cult, and city planning. It has been claimed that Biblical Hebrew is particularly rich in vocabulary related to, for example, farming and water sources. But unlike the Arabic of the Qur'an created in a state of religious fervor, Biblical Hebrew is restrained in its descriptions of and vocabulary for the divine. In short, the corpus is more concerned with a people whose religion was of major importance to it rather than with that religion itself.

Hebrew is not the only language used in the great documents of Judaism. The Palestinian and Babylonian versions of the Talmud (5th to 6th centuries A.D.), excluding the Mishnah (ca. 225 A.D.), are both written mostly in Aramaic, as is the Zohar (late 13th century A.D.). Aramaic is also used for parts of the prayer book and even the Bible itself. Arabic (albeit often written in Hebrew characters) was the language of most of the great works of Jewish theology/philosophy written in territories under Muslim domination, especially Spain. Elsewhere in Europe, Yiddish was employed from the 17th century onward for devotional and ethical literature.

Moreover, even in biblical times, the use of Hebrew for secular (and nonliterary) purposes is attested on hundreds of inscriptions on seals, ostraca, stelae, etc. (These provide a further corpus of 'pre-Rabbinic' Hebrew along with various manuscripts of Ecclesiasticus and the majority of the Dead Sea Scrolls.)

The Decline of Hebrew

After the Babylonian exile, a variety of Hebrew known as 'Rabbinic Hebrew' developed. It existed until the 2nd century A.D. as a popular spoken dialect

(Tannaitic Hebrew), represented in literary form by the Mishnah. Thereafter, it survived another eight centuries as a literary and scholarly dialect (Amoraic Hebrew). However, during even its early phase, Rabbinic Hebrew had to vie with Aramaic and Greek, and it is known that large Jewish communities in Egypt apparently spoke Aramaic (Elephantine) or Greek (Alexandria) exclusively. Although the testimony of Jerome in the late 4th century A.D. indicates a good knowledge of Hebrew among his Jewish informants, the emergence, from the 3rd century B.C. onward, of Greek and Aramaic translations/interpretations of the Bible (*Septuagint*, *Targums*) is a sign of Hebrew's decline. Moreover, because written Hebrew gives relatively few indications of vocalization, there was a danger that unlearned Jews would forget even how to read their Scriptures properly let alone understand them. The situation was made yet more difficult by the existence of versions of the Bible containing different readings of the consonantal text and contentious interpretations of passages by the emerging Christian movement.

The Masoretes

The 6th to 8th centuries saw a flowering of activity among various schools of Masoretes ('bearers of tradition'), who can perhaps be regarded as religiously motivated linguistic theorists. Their aim was to safeguard against corruption of the consonantal text and to provide a system of 'pointing' the basically consonantal Hebrew script to represent how it was to be pronounced at both segmental and suprasegmental levels. One of the systems developed by the Masoretes of Tiberias became dominant and is used in the oldest surviving undamaged manuscript of the whole Hebrew Bible. This document, *Codex Leningradensis B19^A* from 1008–1009, is reproduced in the universally accepted critical edition of the corpus, *Biblia Hebraica Stuttgartensia*. Much earlier texts of parts of the Bible also exist, most notably the second century B.C. Isaiah Scroll from Qumran Cave 1 (unpointed). Texts employing other Masoretic traditions have also survived, and different pronunciations of Hebrew, as well as different trends in morphology, etc., are represented by the various communities of the diaspora (e.g., Ashkenazic, Sephardic, Yemenite).

Hebrew in the Diaspora

After its complete demise as the day-to-day spoken language of the Jewish people and until its 20th-century revitalization as the principal language of the modern State of Israel, Hebrew survived as a language spoken and written by Jews in most diaspora

communities in synagogue worship and religious texts. Hebrew was used, for example, in ceremonial documents, such as Torah scrolls, Passover *haggādōt*, and texts inside phylacteries and *mēzūzōt*, as well as in synagogue and grave inscriptions. With the exception of the sermon and the prayer for the Royal Family (in the UK), Orthodox synagogue services are conducted throughout the world in Hebrew. Elementary Hebrew is traditionally taught to children in a *heder* or 'synagogue-school.' All orthodox Jewish males have to be sufficiently competent in Hebrew to read out loud a portion of Scripture at the age of *bar-mitzvā* (13 years), and, thereafter, when called upon, at ordinary synagogue services. Note as well that Hebrew is imparted within the family in the context of festivals like Passover and *Hanukkah*. At a much more advanced level, Hebrew is also used for instruction within rabbinic academies (*yeshibhōt*).

Literature

The use of Hebrew by Jewish writers never died out, even though its geographical center shifted through time in accordance with the fate of Jewish communities in different countries. Scholars like Rashi, the great 11th-century French Bible commentator, wrote in Hebrew. Indeed, the 'dynamic' of Hebrew – the cause of its internal developments and its ability to adapt to new circumstances, most notably the need to provide a language for what would become the State of Israel – has been one of literature, not speech. In the 12th to 13th centuries A.D., for example, the Ibn Tibbon family, through their translations of Arabic works, accommodated Hebrew to the expression of a wide range of philosophical and scientific topics. Although medieval and earlier literature is frequently religious (including material written at relatively short notice, such as *responsa* to problems arising within particular communities), there is also a wealth of Hebrew poetry, especially from Andalusia, on profane themes.

Secular Contexts

Hebrew was also used as a lingua franca for Jews from different parts of the world, as well as in correspondence and in credit notes, contracts, and other commercial and legal documents within Jewish communities. Records of the Spanish Inquisition attest to the use of Hebrew in oaths, etc., by forced converts to Christianity. From the late 18th century, the widespread use of Hebrew for secular composition developed with the *Haskālā* or Jewish 'Enlightenment.' This had less to do with Hebrew's status as a lingua franca than with the Enlightenment's negative view

of the Yiddish dialects of European Jewry, associated by the reformers with the socially disadvantaged status of Jews and their allegedly low level of cultural achievement. It has also been claimed that Hebrew was in daily use in Palestine during the 19th century.

Hebrew in Other Languages

Jews have usually spoken the dominant language of their host community, on occasions developing specifically Jewish languages/dialects (Yiddish, Ladino, etc.), in which Hebrew has sometimes been of great influence, at least in vocabulary. Where such a development has not taken place, or where a Jewish language/dialect has been superseded by the dominant language, it is in most cases misleading to speak of a Jewish dialect (or sociolect) of a non-Jewish language. But in this situation, Jews continue to use a number of Hebrew expressions for items of Jewish culture (e.g., *siddūr* 'prayer book', *tēphillīn* 'phylacteries', *tallīt* 'prayer-shawl') and in particular contexts (e.g., *mazzāl tōbh* 'congratulations'). The Israeli pronunciation given here represents that used by younger members of Jewish communities, who also tend to use more Hebrew expressions. But the Hebrew pronunciation of older Jews and their vocabulary often reflects that of the Jewish language/dialect once used by themselves or their parents (e.g., Ashkenazic /kōshēr/ for Israeli /kāsh'ēr/ 'kosher', Yiddish *shul* for Hebrew *bēt-kēneset* 'synagogue').

Classical Hebrew has left a few direct traces in the religious vocabulary of other languages, although this has been through the medium of Bible translations rather than that of Jewish communities (e.g., *hallelujah*, *amen*, *behemoth*, *shibboleth*). In the occult, various terms, for example, names for God and other supernatural beings, have been taken over, often in garbled form, from Hebrew. More significantly, the vocabulary and phraseology of the languages of Christian countries have been influenced in a variety of ways by loan-translations from Hebrew via prestigious early, fairly literal, vernacular translations of the Bible. Hebrew also underlies many 'Christian' names (e.g., *Isabel*, *David*, *John*, *Jeremy*, *Sarah*), and is encountered in some Jewish surnames (e.g., *Cohen*, *Levi*, *Rabinowitz*).

The Study of Hebrew within Christianity

Historically, Hebrew has gained the scholarly attention of Christians wanting to gain a better understanding of the Old Testament or to facilitate attempts at conversion of the Jews. Until the beginning of modern

‘scientific’ analysis of the Bible, few non-Jewish scholars could have claimed a familiarity with Hebrew or the ability to contribute to its linguistic analysis on anything approaching the scale of the medieval Jewish grammarians. But it is possible that their efforts aided the long survival of Hebrew, and it is in large measure due to them that Hebrew, elementary Biblical Hebrew at least, still finds a place in the curricula of many universities.

Bibliography

- Aronoff M (1985). ‘Orthography and linguistic theory: The syntactic basis of Masoretic Hebrew punctuation’ *Lg* 61, 28–72.
- Caird G B (1980). *The language and imagery of the Bible*. London: Duckworth.

- Davis M D (1988). *Hebrew deeds of English Jews before 1290*. London: Jewish Chronicle.
- Jones G L (1983). *The discovery of Hebrew in Tudor England: a third language*. Manchester: Manchester University Press.
- Kautzsch E (1910). *Gesenius’ ‘Hebrew Grammar’* (2nd English edn.). Oxford: Clarendon Press.
- Koehler L, Baumgartner W & Stamm J J (1967–1990). *Hebräisches und Aramäisches Lexikon zum Alten Testament* (3rd edn.). Leiden: E. J. Brill.
- Kutscher E Y (1982). *A history of the Hebrew language*. Jerusalem: Magnes Press.
- Parfitt T V (1972). ‘The use of Hebrew in Palestine 1800–1882.’ *Journal of Semitic Studies* 17, 237–252.
- Sáenz-Badillos Á (1993). *History of the Hebrew language*. Cambridge: Cambridge University Press.
- Segal M H (1927). *A grammar of Mishnaic Hebrew*. Oxford: Clarendon Press.

Hebrew, Israeli

G Zuckermann, University of Cambridge, Cambridge, UK

© 2006 Elsevier Ltd. All rights reserved.

Basic Information

The Israeli language (a.k.a. Modern Hebrew) is one of the official languages – with Arabic and English – of the state of Israel, established in 1948. It is spoken to varying degrees of fluency by its 6.8 million citizens (as of September 2004) – as a mother tongue by most Israeli Jews (whose total number slightly exceeds 5 million), and as a second language by Muslims (Arabic speakers), Christians (e.g., Russian and Arabic speakers), Druze (Arabic speakers), and others.

Hebrew was spoken by the Jewish people after the so-called conquest of Israel (c. 13th century B.C.). Following a gradual decline (even Jesus, ‘King of the Jews,’ was a native speaker of Aramaic rather than Hebrew), it ceased to be spoken by the 2nd century A.D. The Bar-Kokhba Revolt against the Romans, which took place in Judaea in A.D. 132–135, marks the symbolic end of the period of spoken Hebrew. For more than 1700 years thereafter, Hebrew was comatose – either a ‘sleeping beauty’ or ‘walking dead.’ It served as a liturgical and literary language and occasionally also as a lingua franca for Jews of the Diaspora, but not as a mother tongue.

Israeli emerged in Eretz Yisrael (or Palestine) at the beginning of the 20th century. Its formation was

facilitated by Eliezer Ben-Yehuda, schoolteachers, and others to further the Zionist cause. Earlier, during the Haskalah (enlightenment) period from the 1770s to the 1880s, writers such as Méndele Mokhér Sfarím (Shalom Abramowitsch) produced works and neologisms which eventually contributed to Israeli. However, it was not until the early 20th century that the language was first spoken.

The genetic classification of Israeli has preoccupied linguists since the language emerged. The traditional school argues that Israeli is Semitic: (Biblical/Mishnaic) Hebrew revived. Educators, scholars, and politicians have contributed to this assumption, linking the history of language to the politics of national revival. The revisionist position, by contrast, defines Israeli as Indo-European: Yiddish relexified, i.e., Yiddish (the revivalists’ mother tongue) is the ‘substratum,’ whilst Hebrew is only a ‘superstratum’ providing the lexis and lexicalized morphology (cf. Horvath and Wexler, 1997). A more recent hypothesis is that Israeli is a hybrid language, both Semitic and Indo-European. It argues that both Hebrew and Yiddish act equally as its primary contributors (rather than ‘substrata’), accompanied by many secondary contributors: Russian, Polish, German, Judeo-Spanish (Ladino), Arabic, English, etc. (see **Figure 1**). Although Israeli phonetics and phonology are primarily Yiddish and its morphology is mainly Hebrew, the European contribution to Israeli is not restricted to particular linguistic domains and is evident even in its morphology.

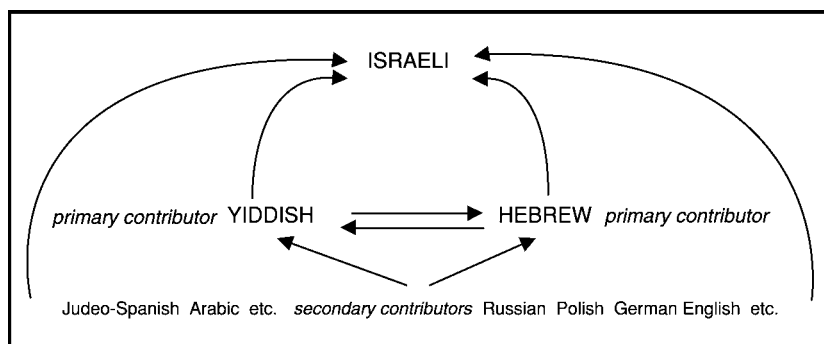


Figure 1 The Israeli hybrid hypothesis.

Thus, the term ‘Israeli’ is far more appropriate than ‘Israeli Hebrew,’ let alone the common signifiers ‘Modern Hebrew’ or ‘Hebrew’ *tout court* (cf. Zuckermann, 1999, 2003, 2005).

Grammatical Profile

Israeli is a fusional synthetic language, with discontinuous, nonconcatenative morphemes with vowel infixation, for example:

yoháv
love.3M.SG.FUT
‘(he) will love’

mitahévet
fall-in-love.3ESG.PRES
‘(she) is falling in love’

yenadvú
volunteer.3PL.FUT
‘(they) will volunteer (others)’

hitnudávti
volunteer.1SG.PAST.COERCIVE/INDUCIVE (*hit-a-é- + -u-á-*)
‘I (was) volunteered (by force)’

However, Israeli is much more analytic than (Biblical/Mishnaic) Hebrew. Whereas the Hebrew phrase for ‘my grandfather’ was *sav-í* ‘grandfather-1.SG.POSS,’ in Israeli it is *sába shel-ì* ‘grandfather GEN-1SG.’ Still, Israeli sometimes uses the Semitic feature known as ‘construct-state’ (Israeli *smikhút*), in which two nouns are combined, the first being modified or possessed by the second. For example, *repúblikat banánot*, literally ‘republic bananas,’ refers to ‘banana republic.’ However, unlike in Hebrew, the construct-state is not highly productive in Israeli. Compare the Hebrew construct-state *?em ha-yéled* ‘mother DEF-child’ with the more analytic Israeli phrase *ha-íma shel ha-yéled* ‘DEF-mother GEN DEF-child,’ both meaning ‘the mother of the child,’ i.e., ‘the child’s mother.’

Israeli is a head-marking language. It is nominative-accusative at the syntactic level and

partially also at the morphological level. As opposed to Biblical Hebrew – whose constituent order is VAO(E)/VS(E) – but like Standard European and English, the usual constituent order of Israeli is AVO(E)/SV(E). Thus, if there is no case marking, one can resort to the constituent order. Israeli is characterized by an asymmetry between definite Os and indefinite Os. There is an accusative marker, *et*, only before a definite O (mostly a definite noun or personal name). *Et-ha* is currently undergoing fusion and reduction to become *ta*. Consider:

| | | | |
|---------------------|------------|-----------|-----------------|
| <i>taví</i> | <i>l-i</i> | <i>et</i> | <i>ha-séfer</i> |
| give.2M.SG.IMP | DAT-1SG | ACC | DEF-book |
| (puristically FUT) | | | |
| ‘give me the book,’ | | | |

where *et*, albeit syntactically a case marker, is a preposition, and *ha* is a definite article. This sentence is realized phonetically as *taví li ta-séfer*.

Israel’s first Prime Minister, David Ben-Gurion, did not like the *et* particle and would have liked to have replaced *taví li et ha-séfer* with *taví li ha-séfer*. (It has been suggested that he was not keen on diplomatic relations with *etyópya* ‘Ethiopia’ for the same reason.) However, such a puristic attitude is hardly ever seen these days and *taví li ha-séfer* is nonnative.

Sound System

Israeli has five vowels: /i, e, a, o, u/. Its consonantal inventory reflects Yiddish (except that the latter has syllabic consonants). Unlike Hebrew, the pharyngealized (emphatic) consonants פ [q], פ [f], and ש [s] have been neutralized and are pronounced [k], [t], and [s] respectively. Hebrew ע [ʕ], נ [ʔ], and ה [h] are all ‘pronounced’ by most Israelis in the same way: most of the time, they are not pronounced. They are only pronounced (both ע and נ as [ʔ], and ה as [h]) when in a postconsonantal position **within uncommon words**. Israeli ה [h] is also pronounced by some speakers at the beginning of phrases. The Hebrew alveolar trill ר [r] is pronounced in Israeli as a unique uvular

approximant [ɣ], similar to the [ʁ] in many Yiddish dialects. For dyslexics, Israeli is much more problematic than Hebrew, the reason being that while Israeli's phonetic system is primarily European, it still uses a phonetically anachronistic Hebrew orthography. Thus, one should not be too surprised to see an Israeli child spelling עקבותי (pronounced *ikvotáv*) 'his traces' as אכווטב.

Whereas the syllable structure of Hebrew was CV(X)(C), that of Israeli is (C)(C)(C)V(C)(C)(C). Israeli does not follow Hebrew spirantization rules. For example, most Israelis say *bekítá bet* rather than the puristic *bekhitá bet* 'in the second grade.' The stress is phonemic, e.g. *bóker* 'morning' and *bokér* 'cowboy.'

Nouns

Israeli nouns show number, normally only singular and plural. Each noun is either masculine or feminine, the latter often being created by adding a suffix to the unmarked masculine. For instance, whereas *mazkír* is 'male secretary,' *mazkirá* is 'female secretary' (note the addition of *-a*). Similarly, whilst *profésor* is 'male professor,' *profésorit* is 'female professor.' Pronouns have 'case forms' consisting of a preposition plus a suffix: nominative (e.g., *aní* 'I'), accusative (*otí* 'me'), dative (*li* 'to me') and genitive (*shelí* 'my'). However, NPs which are not pronouns do not bear case marking. The only exceptions are the above-mentioned accusative marker *et* (or *ta*), and the lexicalized allative ('to/towards') case (which, serendipitously, is based on the historical accusative case), e.g., *báit* 'house' → *ha-báyt-a* 'to the house'; *yerushaláim* 'Jerusalem' → *yerushaláym-a* 'to Jerusalem'; *tsafón* 'north' → *tsafón -a* 'to the north.' New allative phrases, e.g., *tel avív-a* 'to Tel Aviv', are not used unless one is trying to sound flowery or jocular.

Adjectives agree in number, gender, and definiteness with the nouns they modify, e.g.:

| | |
|----------------|----------|
| ha-yéled | ha-gadól |
| DEF-boy | DEF-big |
| 'the big boy'; | |
| yelad-ím | gdol-ím |
| boy-M.PL | big-M.PL |
| 'big boys.' | |

Verbs

As opposed to Biblical Hebrew, which had only a perfect-imperfect distinction, Israeli has three tenses: past, present, and future. In the past and future, verbal forms differ according to gender, number, and person. However, in the present tense, verbs are conjugated only according to gender and number and there is no person distinction. The historical reason

is that the forms of the Israeli present can be traced back to the Hebrew participle, which is less complex than the historical perfect and imperfect forms.

Verbs are transitive, intransitive or ambitransitive (labile). Ambitransitivity is usually of the S=A type, e.g. *dan shatá etmòl* 'Dan_s drank yesterday' (cf. *dan shatá etmòl bíra* 'Dan_s drank yesterday beer_o'). However, owing to Americanization, there are more and more ambitransitive verbs of the S=O type, e.g., *ha-séfer mokhér tov* 'the-book_s sells well' (cf. *grísham mókher et ha-séfer tov* 'Grisham_s sells ACC the-book_o well'); *yésh po máshehu she-meriakh ra* 'There.is here something_s that-smells bad' (cf. *aní meriakh po máshehu ra* 'I_s smell here something_o bad').

Clauses

The main clause in Israeli consists of (a) clause-initial peripheral markers, e.g., discourse markers; (b) NP(s) or complement clause(s); (c) a predicate – either verbal, copular, or verbless; (d) clause-final peripheral elements, e.g. discourse markers. The only obligatory element is the predicate, e.g., *higáti* 'arrive:1SG.PAST.' Sentences (1), (2), and (3) are examples of a verbal, copular, and verbless clause, respectively.

- (1) אסתר אכלה תפוח.
[ester]_A [[akhlá]_V [tapúakh]_O]
[Esther]_A [[eat:3ESG.PAST]_V [apple]_O]
'Esther ate an apple'
- (2) אסתר היא אחות שלי.
[ester]_{CS} [[hi]_{COP} [akhót shel-í]_{CC}]
[Esther]_{CS} [[COP.ESG]_{COP} [sister GEN-1.SG]_{CC}]
'Esther is my sister'
- (3) אסתר חכמה.
[ester]_{VCS} [[khakham-á]_{VCC}]
[Esther]_{VCS} [[clever-F]_{VCC}]
'Esther is clever'

There are many types of subordinate clause, e.g., adverbial (denoting comparison, time, place, condition, concession, reason, result, goal, state), adjectival/relative, nominal/complement. On complementation clauses in Israeli, see Zuckermann (2006).

Concluding Remarks

The grammatical profile of Israeli demonstrates its binary nature, which has important theoretical implications for many branches of language science: contact linguistics, sociolinguistics, language revival/survival, linguistic genetics and typology, creolistics, and mixed languages. Genetic affiliation – at least in the case of (semi-)engineered, 'nongenetic' languages – is not discrete but rather a continuous line. The comparative method and lexicostatistics, though

elsewhere useful, are not here sufficient. Linguists who seek to apply the lessons of Israeli to the revival of no-longer spoken languages should take warning.

Israeli affords insights into the politics not only of language, but also of linguistics. One of the practical implications is that universities, as well as Israeli secondary schools, should employ a clear-cut distinction between Israeli linguistics and Hebrew linguistics. Israeli children should not be indoctrinated to believe that they speak the language of Isaiah – unless the teacher is referring to the 20th-century Israeli polymath and visionary Isaiah Leibowitz. Although revivalists have engaged in a campaign for linguistic purity, the language they created often mirrors the very cultural differences they sought to erase. The study of Israeli offers a unique insight into the dynamics between language and culture in general and in particular into the role of language as a source of collective self-perception.

Bibliography

- Horvath J & Wexler P (eds.) (1997). *Relexification in creole and non-creole languages, with special attention to Haitian Creole, Modern Hebrew, Romani, and Rumanian*. Wiesbaden: Otto Harrassowitz.
- Zuckermann G (1999). 'Review of Nakdimon Shabbethay Doniach and Ahuvia Kahane (eds.), *The Oxford English-Hebrew Dictionary*, Oxford University Press, 1998.' *International Journal of Lexicography* 12, 325–346.
- Zuckermann G (2003). *Language contact and lexical enrichment in Israeli Hebrew*. London/New York: Palgrave Macmillan.
- Zuckermann G (2005). *Hebrew as a myth*. Tel Aviv: Am Oved.
- Zuckermann G (forthcoming). 'Complementation in Israeli.' In Aikhenvald A Y & Dixon R M W (eds.) *Complementation: a cross-linguistic typology*. Oxford: Oxford University Press.

Highland East Cushitic Languages

G Hudson, Michigan State University, East Lansing, MI, USA

© 2006 Elsevier Ltd. All rights reserved.

Demography

Highland East Cushitic (HEC) languages are spoken by some five and a half million people in south-central Ethiopia, in an area bounded generally by the 5th and 9th degrees of north latitude and the 37th and 39th degrees of longitude.

There are five HEC languages, from north to south Hadiyya, Kambaata, Sidaama, Gedeo (formerly Darasa), and Burji (formerly sometimes Bambala). A dialect of Hadiyya (i.e., mutually intelligible with it) separately reported in the 1994 census is Marak'o (or Libido). Dialects of Kambaata also separately reported in the census are T'imbaaro and a closely related pair Alaba and K'abeena. (Letters followed by apostrophes represent glottalic ejective consonants.) Sidaama has distinct regional dialects, as do Gedeo and Burji, but these are not usually distinguished by name or otherwise in the literature. A group of Burji left Ethiopia around the turn of the century, and live in and about the northern Kenya town of Marsabit.

Table 1 shows the numbers of ethnic-group members and first-language speakers of the five languages and nine named dialects as reported by the 1994

Ethiopian Census (Office of Population and Housing Census Commission of Ethiopia, 1998, vol. 1; also see Hudson, 2003). The 1994 Census reported 4 241 804 HEC first-language speakers, which is 7.9% of the 1994 Ethiopian total population of 53 477 265. Sidaama, the most populous HEC language, was the fifth most populous Ethiopian language, after Amharic (17 372 913), Oromo (16 777 975), Tigrinya (3 224 875), and Somali (3 187 053). The Census (Summary Report, 1998) suggested a multiplier of 1.37 to estimate 2005 total

Table 1 HEC ethnic-group members and first-language speakers

| HEC dialect | Ethnic-group members, 1994 census | First-language speakers, 1994 census |
|-------------|-----------------------------------|--------------------------------------|
| 1 Burji | 46 552 | 35 731 |
| 2 Gedeo | 639 879 | 637 082 |
| 3 Hadiyya | 927 747 | 923 957 |
| 4 Marak'o | 38 093 | 36 612 |
| Total 3–4 | 965 840 | 960 569 |
| 5 Kambaata | 499 631 | 487 654 |
| 6 Alaba | 125 894 | 126 257 |
| 7 K'abeena | 35 065 | 35 783 |
| 8 T'imbaaro | 86 499 | 82 803 |
| Total 5–8 | 747 089 | 732 497 |
| 9 Sidaama | 1 842 444 | 1 876 329 |
| Total 1–9 | 4 241 804 | 4 242 208 |

Ethiopian population, which yields a 2005 HEC first-language-speaker total of over 5.8 million. In fact, although Hadiyya and Sidaama may have increased speakers in such proportion, less populous varieties may have decreased.

Notice in **Table 1** that Alaba, K'abeena, and especially Sidaama are reported to have more first-language speakers than ethnic-group members. For Sidaama, this is consistent with its large number of speakers and presumptive function as a lingua franca in its region. The reason for this result for Alaba and K'abeena is less apparent.

The HEC territory is part of a linguistically diverse region of the West Rift Valley highlands where converge three of the six subgroups of Afroasiatic (Hamito-Semitic) languages: Cushitic, Semitic, and Omotic (the other three are Egyptian, Berber, and Chadic). Most of the HEC peoples share with Semitic 'Gurage' and Omotic peoples of this region a unique agro-ecology based on cultivation of the 'false banana' tree *ensete adulis*.

Classification

The HEC languages are Afroasiatic languages of the Eastern branch of Cushitic. The majority of Cushitic languages, some 19, are spoken in Ethiopia. The most populous Eastern Cushitic languages are Oromo and Somali. For an overview of Afroasiatic linguistics, see Hayward (2000). For a study of Cushitic grammatical characteristics, see Hetzron (1980), and for an overview of Cushitic classification, see Tosco (2000).

Figure 1 presents the tree-diagram of subgroupings within HEC, assuming mutual intelligibility within the Kambaata group. Burji is considerably divergent from the others, presumably reflecting its earlier separation from them. The general picture based on linguistic diversity, which is focused in southwest Ethiopia, is of south to north spread, with Burji the least moved and Hadiyya having diverged most recently from Kambaata. Oral traditions of these people often claim northern origins, but this

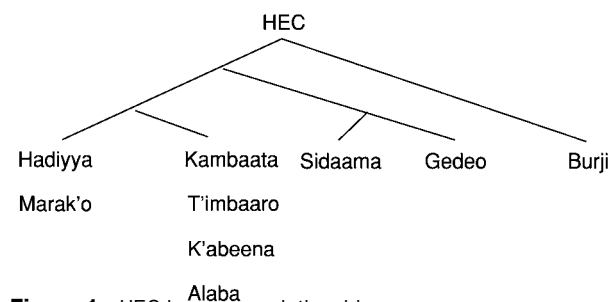


Figure 1 HEC language relationships.

is probably an influence of their Christian and/or Muslim faith.

Writing

The HEC languages were little written until the linguistic liberalization resulting after the Ethiopian revolution of 1974, which ended the unique official status of Amharic, and made Gedeo, Hadiyya, Kambaata, and Sidaama among the 15 languages promoted for literacy by the new government. Then the languages were written in the Amharic (Ethiopic) writing system. More recently, Ethiopian Cushitic languages have begun commonly to be written with the European-language alphabet. With further linguistic liberalization in the 1990s, Sidaama and Hadiyya, as the most populous languages of their areas, are now used, if limitedly, in primary education and local government, and in publications including newspapers, political writings, and fiction.

Typology

The HEC languages are typologically quite consistent as inflectional, suffixing, and head-final ('SOV'), and have closed syllables of limited types. A measure suggestive of diversity within HEC is percentages of cognates in a basic hundred-word list. **Table 2** presents these figures for the five major HEC varieties, which range from 39% for Burji-Kambaata to 70% for Gedeo-Sidaama (Wedekind, 1990: 46). Mutual intelligibility between varieties is roughly expected by such figures from about 75–80% (see also Bender and Cooper, 1971).

A broad descriptive-comparative survey of the HEC languages can be found in Hudson (1976); additional comparative information focused on morphophonemics in Abebe *et al.* (1985); a study of subclassification and history in Hudson (1981); a Burji etymological dictionary in Sasse (1982); an HEC comparative dictionary in Hudson (1989); a comparison of Kambaata varieties in Crass (2001); and a survey of HEC morphology in Hudson (2007). Analysis of Burji, Gedeo, and Sidaama texts is presented by Wedekind (1990) and of a Burji text by

Table 2 Percentage of cognates shared by five HEC varieties in a basic 100-word list

| | Burji | Gedeo | Sidaama | Kambaata |
|----------|-------|-------|---------|----------|
| Hadiyya | 44 | 56 | 62 | 66 |
| Kambaata | 39 | 49 | 66 | |
| Sidamo | 47 | 70 | | |
| Gedeo | 43 | | | |

Kellner (2001). Some characteristic features of HEC grammar concerning phonology, verb and noun morphology, and syntax are discussed next.

Phonology

Typical HEC word structure, as well as information about other phonological characteristics, is suggested in Table 3, a selection of 10 basic words in the 5 languages. Words of Table 3 are written phonemically, but phonetic interpretation is straightforward. Characteristics that may be noted are typically open syllables, five-vowel system, contrastive vowel and consonant length, glottalic ejective consonants *t'*, *k'*, *č'* (rarely also *p'*), glottalic implosive *d'* (in Burji 'bird', 'ear'), phonemic glottal stop, glottal onset *ʔm* ('bite'; there are also *ʔn* and *ʔl*), and word-internal syllables closed by sonorant consonants or the onset of long-consonants. In Sidaama and Kambaata, there are cases of final-syllable stress contrast, apparently resulting from shortened long vowels (cf. Sidaama and Gedeo 'blood').

Regarding HEC subclassification (Figure 1), apparent in Table 3 is the divergence of Burji, with its lesser number of apparent shared cognates, and the validity of subgroups Hadiyya-Kambaata and Gedeo-Sidaama, within which are more apparent cognates.

Characteristic of HEC languages other than Burji are pervasive processes of *i*-epenthesis, consonant

assimilation, and nasal metathesis that 'conspire' to assure allowed syllable contacts when stems and suffixes combine in verb formation. These processes are illustrated in Table 4, which shows Gedeo past-tense suffixes in combination with verb stems *af*- 'get', *dar*- 'tear', and *aff*- 'perspire', the first with progressive assimilation by its stem-final obstruent of *t* of *t*-initial suffixes, the second with progressive assimilation by its stem-final sonorant consonant of *n* of *n*-initial suffixes, and the third with *i*-epenthesis between its stem-final long consonant and the consonant-initial suffixes.

Verb Morphology

The Gedeo examples of Table 4 also exemplify the typical paradigm of HEC verb formation, in which a monosyllabic verb stem combines with suffixes whose initial part is cognate with those of a general Afroasiatic pattern (seen as subject-prefixes in Semitic and Berber): first singular with an initial vowel (Gedeo *-enne*), second singular, second plural and third singular feminine in *-t* (*-titto*, *-ti-ne*, and *-te*), and plurals in *-n* (*-n-enne*, *-ti-ne*, *-ne*). Sidaama adds to the paradigm gender agreement in first-person and second-person singular, with final *o* for masculine and *a* for feminine, e.g., *afummo* 'I-masc. got', *afumma* 'I-fem. got'.

Also typical of HEC, with cognates elsewhere in Cushitic, are regular verb derivatives for causative in *-is*, passive in *-am*, and reflexive in *-id'* and its reflexes. Some examples from Sidaama are *āja* 'decrease' (vi.), causative *āja* (vt.); *k'ana* 'suck', causative *k'ansa* 'suckle', *fana* 'open', passive *fanama*; *duna* 'pour', passive *dunama* 'leak', *afa* 'get', reflexive *afir'a*, *hasa* 'seek', reflexive *hasir'a*. Sidaama *r* (of *-ir' < -id'*) is distinct from plain *r* only as it glottalizes preceding consonants, e.g., *k'ʔna* 'suckle,' reflexive of *k'ana* 'suck'.

A peculiar lexical phenomenon of Ethiopian languages whether Cushitic, Semitic, or Omotic, but perhaps particularly common in HEC, is verb compounds formed by words peculiar to the idioms plus the verbs 'say' for intransitives and 'do' for

Table 3 Ten basic words in five HEC varieties

| Gloss | Burji | Gedeo | Hadiyya | Kambaata | Sidaama |
|----------|----------|--------|---------------|------------|----------------|
| bird | č'iid'aa | č'iʔa | č'iʔ- iččo | č'iičču-ta | č'eeʔ- iččo |
| to bite | gama | gaʔma | gaʔmi | gaʔmu | gaʔma |
| bone | mič'a | mičč'o | mik'e | mik'a | mik'a |
| blood | č'eeji | mundee | t'iiga | k'egu | mundé |
| breast | ununa | unuuna | anuuna | anuuna | unuuna |
| to come | inta(j)a | daga | waari | waalu | daga |
| to die | re(j)a | re(ʔ)a | lehi | re(e)hu | rea |
| ear | d'aga | manša | mačč'e | mačč'a | mačč'a |
| five | umutta | onde | onto | onto | onte |
| foot/leg | luka | lekka | lokko | lokka-ta | lekka |

Table 4 Epenthesis, assimilation, and metathesis in Gedeo past-tense verb formation

| | Suffix of past tense | <i>af</i> - 'get' | <i>dar</i> - 'tear' (vt) | <i>daff</i> - 'perspire' |
|---------|----------------------|-------------------|--------------------------|--------------------------|
| 1 sg. | -enne | afenne | darenne | daffenne |
| 2 sg. | -titto | affitto | dartitto | daffitto |
| 3 sg.m. | -e | afe | dare | daffe |
| 3 sg.f. | -te | affe | darte | daffite |
| 1 pl. | -nenne | a[m]fenne | darrenne | daffinnenne |
| 2 pl. | -tine | affine | dartine | daffitine |
| 3 pl. | -ne | a[m]fe | darre | daffine |

Table 5 Representative HEC singulars and plurals

| Burji | Gedeo | Hadiyya | Kambaata | Sidaama |
|---|---|--|---|---|
| gota (sg.), gotanno 'sheep' | dureessa (sg.), dureeyye 'rich ones' | gaamelakiččo, (sg.), gaamela 'camels' | aburču (sg.), aburrata 'roosters' | hoončo (sg.), hoonna 'juniper trees' |
| hiliččo (sg.), hilaano 'calves' | geerčo (sg.), geeŕre 'old men' | hamašiččo (sg.), hamašša 'snakes' | bezzeečču (sg.), bezzeebeezaa 'stars' | ille (sg.), illubba 'eyes' |
| č'uuwe (sg.), č'uuweenna 'chickens' | reččo (sg.), reŕe 'goats' | kina (sg.), kinnewwa 'stones' | lokkata (sg.), lokaakkata 'feet, legs' | ibiččo (sg.), ibiibe 'lice' |

Table 6 Two sentences in HEC languages

| Gloss | HEC language | Sentence |
|-------------------------------------|--------------|--|
| 'This cow is fat' | Hadiyya | tu saay dilba-tte <i>this.fem cow fat-be.fem</i> |
| | Kambaata | ku saŕo gaana <i>this.masc cow fat</i> |
| | Sidaama | tini saa lowo-te <i>this cow fat/big-be.fem</i> |
| | Gedeo | tinni saa-yy-iččo furda-tt'e <i>this.fem cow-nom-sing fat-be.fem</i> |
| 'I cut (past) meat with this knife' | Burji | ku saay-i gabboo-na <i>this.masc cow-nom fat-be</i> |
| | Hadiyya | maar ka billawi-n mur-ummo <i>meat this.masc knife-with cut-I.past</i> |
| | Kambaata | maala ka-n billawi-n murr-oommi <i>meat this.masc-with knife-with cut-I.past</i> |
| | Sidaama | maala tenné seet'e-nni mur-umm-o <i>meat this.fem knife-with cut-I.past-masc</i> |
| | Gedeo | maala konne šiifi-nni kut-enne <i>meat this.masc knife-with cut-I.past</i> |
| | Burji | maala ta sore-čč-ina mur-anni <i>meat this.fem knife-with-focus cut-I.past</i> |

transitives. Examples with 'say' are Burji *naač'i i* 'smile', Gedeo *dapp'i hiyy-* 'be tight', 'be quiet', Hadiyya *heešš y-* 'stoop', Kambaata *abb y-* 'rise', and Sidaama *beebi y-* 'be quiet'.

Noun Morphology

HEC has a singular ('singulative') as well as plural suffix, the former basically -č, or -iččo after obstruents. Plural formations are various, sometimes involving internal change, which is perhaps a Cushitic

or Afroasiatic characteristic. Table 5 presents representative examples for all five languages.

The languages generally mark nominatives (and/or definites) with -i, but case marking is complex, perhaps under change, and more research is needed on this and others aspects of HEC grammar. Nominative vs. accusative marking has sometimes been considered a postergative phenomenon, as a generalization of the 'ergative' or subject-of-transitive case.

Sidaama (and perhaps other HEC languages to a lesser extent) has a phenomenon of gender-exclusive vocabulary (Anbessa, 1987). By a taboo on words beginning with the first syllable of her father-in-law's name, a woman must circumlocute or substitute words fixed in women's language for this purpose. For example, a woman whose father-in-law's name begins with *ma* would use the word *basara* 'meat' instead of the usual *maala*.

Syntax

HEC syntax may be best presented in brief as in Table 6, which compares in the five HEC languages two sentences, one copular and one with a transitive verb.

Bibliography

- Abebe Gebre-Tsadik M, Sim R, Wedekind C & Wedekind K (1985). 'The morphophonemics of five Highland East Cushitic languages including Burji.' *Afrikanistische arbeitspapiere* 2.
- Anbessa Teferra (1987). 'Ballišša: women's speech among the Sidama.' *Journal of Ethiopian Studies* 20, 44–59.
- Bender M L & Cooper R L (1974). 'Mutual intelligibility within Sidamo.' *Lingua* 27, 32–52.
- Crass J (2001). 'The position of K'abeena within Highland East Cushitic.' *Afrikanistische arbeitspapiere* 67, 5–60.
- Hudson G (1976). 'Highland East Cushitic.' In Bender M L (ed.) *The non-Semitic languages of Ethiopia*. East Lansing: African Studies Center, Michigan State University. 232–277.
- Hudson G (1981). 'The Highland East Cushitic family vine.' *Sprache und geschichte in Afrika* 3, 97–124.

- Hudson G (1989). *Highland East Cushitic dictionary* (Kuschitische Sprachstudien 7). Hamburg: Buske.
- Hudson G (2003). 'Linguistic analysis of the 1994 Ethiopian census.' *Northeast African Studies* 6(3), 89–107.
- Hudson, G. (2007). 'Highland East Cushitic morphology.' In *Morphologies of Asian and African languages*, Alan Kaye (ed.), Winona Park: Eisenbrauns.
- Hayward R J (2000). 'Afroasiatic.' In Heine B & Nurse D (eds.) *African languages: an introduction*. Cambridge: Cambridge University Press. 74–98.
- Kellner A (2001). 'The lazy baboon: a linguistic analysis and anthropological interpretation of a Burji tale.' *Afrikanistische Arbeitspapiere* 65, 41–69.
- Office of Population and Housing Census Commission. (1998). *1994 Population and housing census of Ethiopia: results at country level* (Vol. 1). Addis Ababa: Central Statistical Office.
- Office of Population and Housing Census Commission. (1998). *1994 Population and housing census of Ethiopia:*

results at country level, summary report. Addis Ababa: Central Statistical Office.

Sasse H-J (1982). *Etymological dictionary of Burji* (Kuschitische Sprachstudien 1). Hamburg: Buske.

Tosco M (2000). 'Cushitic overview.' *Journal of Ethiopian Studies* 33(2), 87–121.

Wedekind K (1990). *Generating narratives: interrelationships of knowledge, text variants, and Cushitic focus strategies*. Berlin: Mouton de Gruyter.

Relevant Websites

<http://www.sidaamaconcern.com> – See this website for Sidaama writing.

<http://www.msu.edu/~hudsonHECrefs.htm> – A full bibliography of HEC languages can be found here.

Hiligaynon

R D Zorc, Language Research Center, McNeil Technologies Inc., Hyattsville, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Hiligaynon is the fourth largest language of the Philippines, representing approximately 10% of the national population. Its seven million speakers are located throughout Negros Occidental, southeastern Panay, Guimaras Island, and in urban centers of Mindanao (Davao and Zamboanga) and of Palawan (Puerto Princesa). It is a major trade language of the

Western Visayan region (e.g., Antique and Aklan). Ilonggo, its alternate name, originally specified the dialect of Iloilo. It has many dialects with minor variations from town to town. The most distinct are Capiznon (Capiz Province) and Kawayan (south of Bacolod City). It is a member of the Central Bisayan subgroup along with Waray, Masbateño, and Romblomanon. These are, in turn, members of the Bisayan group of Central Philippine languages, including Tagalog and Bikol (Zorc, 1977), all of which are ultimately descended from Proto-Austronesian.

Although legends and fabrications abound (see Scott, 1984), nothing is known historically prior to the Spanish. Alzina recorded that the Hiligaynons of Oton (Panay) traced their origin to Leyte (Kobak, 1969–70: 22), which correlates with the subgrouping.

Table 1 Hiligaynon sound system

| Consonants | | | | |
|-------------|------------|------------------|------------|----------|
| Stops | Labial | Apical | Velar | Glottal |
| voiced | b | d | g | |
| voiceless | p | t | k | ' |
| Fricatives | (f) | s | | h |
| Affricates | | | | |
| voiced | | (j) [dy] | | |
| voiceless | | (ch) [ts] | | |
| Continuants | | | | |
| liquid | | l | | |
| rhotic | | r | | |
| semivowel | w | y | | |
| Nasals | m | n | ng | |
| Vowels | | | | |
| | Front | Central | Back | |
| High | i | | u | |
| Mid | (e) | | (o) | |
| Low | | a | | |

Table 2 Hiligaynon Pronouns

| Pronoun | Topic | Oblique forms | | Locative |
|------------------|------------------|---------------|--------------------|----------------|
| | | preposed | postposed | |
| I | akó | ákon | -ko / nákon | sa'ákon |
| you [singular] | ikáw / ka | ímo | -mo / nímo | sa'ímo |
| he/she | siyá | íya | níya | sa'íya |
| we [+you / incl] | kitá | áton | -ta / náton | sa'áton |
| we [-you / excl] | kamí | ámon | námon | sa'ámon |
| you [plural] | kamó | ínyo | nínyo | sa'ínyo |
| they | silá | íla | níla | sa'íla |

The basic phonology of Hiligaynon consists of 16 consonants and 3 vowels; accent (stress) is contrastive. Native speakers educated in Spanish and English have an additional three consonants /f, j, ch/ and two vowels /e, o/. Accent (/á, í, ú/ with vowel length) occurs in an open penult. The vowel [o] is an allophone of /u/ in final syllables but is phonemic in loans. Accent predictably falls on a closed penult: *táytay* ‘bridge.’

The glottal stop is written as a hyphen when it appears before another consonant: *bág-o* ‘new,’ *búg-at* ‘heavy,’ *gáb-i* ‘evening.’ It is ignored word-finally in most local publications; linguists have spelled it with *q* or an apostrophe. Accent, which is also not represented in the orthography, is critical in distinguishing words or derivations:

| | |
|------------------------------|-----------------------------------|
| <i>ámo</i> ‘boss’ {Spanish} | <i>amó</i> ‘thus, like that’ |
| <i>sá’og</i> ‘crawl’ | <i>sa’óg</i> ‘wear out by use’ |
| <i>bílin</i> ‘remain, stay’ | <i>bilín</i> ‘leftovers’ |
| <i>pikot</i> ‘mend’ | <i>pikót</i> ‘half-closed (eyes)’ |
| <i>lútu</i> ‘to cook’ | <i>lutú</i> ‘cooked’ |
| <i>túbo</i> ‘pipe’ {Spanish} | <i>tubó</i> ‘sugarcane’ |

Various morphophonemic changes apply in inflection and derivation:

Table 3 Hiligaynon deictics

| | <i>Near me</i> | <i>Near you</i> | <i>Far away</i> |
|-------------|----------------|-----------------|-----------------|
| Topic | iní | iná’ | ató |
| Oblique | siní | siná’ | sádto |
| Locative | dirí | dirá’ | dídto |
| Existential | yári | yára’ | yádto |
| Verbal | karí | kará’ | kádto |

Table 4 Hiligaynon verb inflection

| <i>Verbs</i> | <i>Past</i> | <i>Progressive</i> | <i>Contingent</i> | <i>Future</i> | <i>Command</i> |
|---------------|--------------------|---------------------|-------------------|------------------|----------------|
| Active | | | | | |
| Punctual | -um- | | -um- | ma- | mag- |
| Durative | nag- | naga- | mag- | maga- | pag- |
| Distributive | naN- | nagapaN- | maN- | magapaN- | magpaN- |
| Potential | naka- | naka- | maka- | maka- | |
| Passive | | | | | |
| Punctual | -in- | | -(h)on | -(h)on | -a |
| Durative | gin- | gina- | pag – on | paga – on | pag – a |
| Distributive | ginpaN- | ginapaN- | paN – on | paN – on | |
| Potential | na- | na- | ma- | ma- | |
| Instrumental | | | | | |
| Punctual | -in- | | i- | i- | i- |
| Durative | gin- | gina- | i(g)- | iga- | ipag- |
| Distributive | ginpaN – an | ginapaN – an | ipaN – | ipaN- | |
| Potential | (ki)na- | na- | ika- | ika- | |
| Local Passive | | | | | |
| Punctual | -in – an | | -an | -an | -i |
| Durative | gin – an | gina – an | pag – an | paga – an | pag – i |
| Distributive | ginpaN – an | ginapaN – an | paN – an | paN – an | |
| Potential | na – an | na – an | ma – an | ma – an | |

Intervocalic /d/ > [-r-]: *báyad* ‘pay’ > *bayáran* ‘be paid’

With Spanish verbs, final /r/ changes to [-h-]: *probár* ‘try’ > *probahán* ‘be tried.’ Nasal final prefixes such as the distributive pang- yield nasal assimilation and consonant loss:

batí ‘hear’ > *památi* ‘listen to,’ *tíndog* ‘stand’ > *panindúgan* ‘position,’ *káhoy* ‘wood’ > *pangahóy* ‘gather firewood’

Vowel loss is common with suffixation:

inóm ‘drink’ > *ímnon* ‘be drunk,’ *sunúd* ‘follow’ > *súndun* ‘be followed’

Grammatical relations are shown by particles (*kag* ‘and,’ *na* ‘now, already,’ *mga* plural, *man* ‘also, too,’ *lang* ‘only’) or affixes: prefixes (*pag-* temporal verb, *ka-* companion noun), infixes (*-in-* passive past), suffixes (*-un* direct passive, *-an* local passive), or circumfixes (*ka* – an abstract noun, *gina* – an local passive progressive).

Nominals are inflected for case: common nouns (marked by *ang* topic, *singin* indefinite oblique, *sang* definite oblique, *sa* locative) or personal names (*si* topic, *ni* oblique, *kay* locative; plural: *sanday* topic, *nanday* oblique, *kanday* locative). Demonstratives orient to person, locus, time, or anaphora. They have existential and verbal inflections.

Verbs are inflected for four voices (active, passive, instrumental, local), four tenses (past, progressive, contingent, future), three aspects (punctual,

durative, distributive), and three moods (factual, command, potential).

Unmarked word order is V-S-O (verb-subject-object); because nominal constituents are case marked, word order can be free. Initial position by any nonverb usually serves to highlight or contrast.

- (1) kahápon si Hwaníng nag'abút
yesterday TOP Johnny past active-arrive
'It was yesterday that Johnny arrived'

Two other markers are: *nga* (-*ng* after vowels), a ligature uniting nouns with other constituents and *ka* after numerals.

- (2) matahúm nga babáyí
pretty LINK woman
'a pretty lady'
- (3) ma'áyo-ng ága
ADJ-good-LINK morning
'good morning'
- (4) duhá ka simána
two NUM week
'two weeks'

There are three negatives: *ayáw* 'don't!' IMPERATIVE, *walá'* + TOPIC or *waláy* + OBJ 'none' EXISTENTIAL, 'did not' PAST or 'doesn't' PRESENT, and *díli'* 'will not' FUTURE or PREDICATIVE 'is not so'; a fourth, *bukún*, negates nouns and adjectives in some dialect areas.

- (5) Walá' kitá sing baláy
NEG-EXIS we (incl) OBL house
'We have no house'
- (6) Waláy baláy kitá
NEG-EXIS house we (incl)
'We have no house'
- (7) díli' siyá manggaránun
bukún siyá manggaránun
neg-pred he/she rich
'He is not rich'

Bibliography

- Constantino E (1975). *An English-Hiligaynon dictionary*. Quezon City: University of the Philippines.
- Kaufmann, J. (n.d.) *Visayan-English dictionary* (*Kapulungan Binisayá-Ininglís*). Iloilo: La Editorial.
- Kaufmann J (1939). *Principles of Visayan grammar*. Manila: Catholic Trade School.
- Kobak C (1969–70). 'Alzina's historia de las Islas E Indios de Bisayas.' *Leyte-Samar Studies* Chapters 1–3, 3/1, 14–36, Chapters 4–5, 4/1: 17–28.
- McFarland C D (1980). *A linguistic atlas of the Philippines*. Study of Languages and Cultures of Asia and Africa Monograph Series, No. 15. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Motus C (1971). *Hiligaynon dictionary*. Philippines: PALI Language Texts. Honolulu: University of Hawaii Press.
- Motus C (1971). *Hiligaynon lessons*. Philippines: PALI Language Texts. Honolulu: University of Hawaii Press.
- Scott W H (1984). *Prehispanic source materials for the study of Philippine history* (rev. edn.). Quezon City: New Day Publishers.
- Sunio D & Zorc R D (1992). *Hiligaynon reader*. Kensington, MD: Dunwoody Press.
- Wolfenden E P (1971). *Hiligaynon reference grammar*. Philippines: PALI Language Texts. Honolulu: University of Hawaii Press.
- Wolfenden E P (1975). *A description of Hiligaynon syntax*. Summer Institute of Linguistics Publications in Linguistics and Related Fields, Publication Number 46. Norman: University of Oklahoma Press.
- Yap J E (1983). *English-Visayan. (Ilonggo) dictionary* (*Kapulungan Iningles-Binisaya*). Iloilo: Apostol Printers Inc.
- Zorc R D (1967). *Peace Corps primer for the Western Visayas*. Washington: Government Printing Office.
- Zorc R D (1977). *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Pacific Linguistics C.44. Canberra: The Australian National University.
- Zorc R D (2001). 'Hiligaynon.' In Garry J & Rubino C (eds.) *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*. New York: New England Publishing Associates. 300–304.

Hindi

S Shukla, Georgetown University, Washington, DC, USA

© 2006 Elsevier Ltd. All rights reserved.

The Indo-Europeans came to India from the north-west about 4000 years ago. In India they called themselves *Arya*, 'noble honorable,' and called their country *Aryavarta*, 'abode of the noble.' Because of

this ancient reference, today their language is known as *Indo-Aryan*. In regard to its structure and its continuity, Indo-Aryan can be divided into three periods: Old Indo-Aryan, Middle Indo-Aryan, and Modern Indo-Aryan.

Old Indo-Aryan is mostly represented by Sanskrit, which includes both Vedic and Classical Sanskrit. Middle Indo-Aryan is represented by three successive stages of development: Pali, Prakrit, and Apabhramsha.

Pali is the language of the canonical writings of the Theravada school of Buddhism. The various dialects recorded in the inscriptions of Ashoka (c. 250 B.C.) and other early inscriptions also belong to this first period. Prakrit, found mainly in drama and in the religious writings of the Jains, represents the second stage of development. Apabhramsha, though known from texts of the 10th century A.D., was undoubtedly formed prior to this date. It represents the final stage of Middle Indo-Aryan. Hindi and modern Indo-Aryan languages, such as Assamese, Bengali, Gujarati, Kashmiri, Marathi, Nepali, Oriya, Panjabi, Sindhi, Sinhalese, and others, can be dated to about the end of the 10th century A.D. From then on, their development shows a gradual transformation into their present form.

As an Indo-Aryan language, Hindi is a branch of the Indo-European family of languages, and thus is a distant cousin of English, French, Greek, Russian, Spanish, and other Indo-European languages. The name 'Hindi' is a Persian word referring to the people who lived in the Sindhu river area. Later the word was used as a name for the language spoken around Delhi. This language has been called by other names, such as Hindavi, Hindui, and Hindustani. Hindi and Urdu are variants of the same language, which, in its common spoken form, used to be called Hindustani. Hindi is written in the Devanagari script, derived from one of the scripts used to write Sanskrit, while Urdu is written in a modified version of the Persian script, itself derived from the Arabic script. Along with Urdu, Hindi has been the dominant language of modern India and has had an impact on other Aryan and non-Aryan languages spoken in the country. Today it is spoken in most of India. In terms of total number of speakers, it ranks third in the world after Chinese and English. The percentage of the population of India that speaks Hindi is growing, and ranges upwards from 45%. Large language communities outside of India, including Nepal, Pakistan, Singapore, Malaysia, Burma, Mauritius, Trinidad, Guyana, and several countries in eastern and southern Africa also speak Hindi. Furthermore, Hindi is taught at many universities in the United States, Russia, Britain, and the Near East, as well as in other parts of Asia. Today Hindi is a symbol of Indian unity and nationality. It is the national language of India and the official state language of Bihar, Delhi, Haryana, Himachal Pradesh, Madhya Pradesh, Rajasthan, and Uttar Pradesh. Since Hindi has the largest number of speakers of any language in India, it is the medium of a great number of political, social, and cultural activities. Consequently, the economic and political influence of Hindi in India cannot be overlooked.

There are regional differences in Hindi, affected by the other languages that people speak. Within each regional variety of Hindi, there is considerable variation in speech according to the education and social standing of the speakers. This creates social dialects. Naturally, educated speech has more prestige and is thus embraced by government agencies, learned professions, political parties, the media, and institutions that attempt to communicate beyond their regional boundaries. It is this educated Hindi that has acquired the status of standard Hindi.

Underlying all these varieties, however, is a nucleus or common core shared by all speakers of Hindi. Unlike the other major modern Indo-Aryan and Dravidian languages of India, Hindi is not exclusively associated with any one region or province. Although it is the home language of a relatively small number of speakers (in Haryana, western Uttar Pradesh, north-eastern Madhya Pradesh, and portions of eastern Rajasthan), Hindi in its various forms is a spoken and written language of practically all of northern India. It is also the most commonly understood Aryan language in the Dravidian south. Associated with all the different varieties of Hindi are the forms of formal and informal Hindi. The language used when giving instructions to a construction worker varies from the language used when discussing politics or poetry. Typically, the switch involves utilizing a particular set of lexical items habitually used for handling the topic in question. This tendency has produced informal Hindi, the language of everyday speech, which includes instructions to servants, waiters, workmen, clerks, etc., and formal Hindi, the language of films, newspapers, news magazines, education, and literature. Formal Hindi is often marked by an abundance of words borrowed from Sanskrit. When the majority of borrowing is from Persian and Arabic sources instead of from Sanskrit, the language becomes formal Urdu.

Informal Hindi is usually used in speaking to children, and thus is acquired by children as their mother tongue. Children are often exposed to formal Hindi through their parents, but the actual learning of this variety is accomplished through formal education. In areas where the native language is different from Hindi, most speakers acquire Hindi through formal education. As stated earlier, Hindi spoken in these areas bears the influence of the native language, including elements from the native lexicon and phonology.

The sound systems of formal and informal Hindi constitute a single phonological system shared by both varieties; there are only a few phonemes and phonological rules found exclusively in formal or in informal Hindi. Like any language, the lexicon of

Hindi consists of native words that have organically evolved from an earlier Indo-Aryan form and non-native words that have been borrowed from other languages. Sanskrit has been the primary source for borrowing in Hindi. After Sanskrit come Persian, Arabic, English, Turkish, and Portuguese borrowings, in that order. There are also a number of Dravidian words in Hindi, but most of them have come through Sanskrit. Hindi has also borrowed words from other modern Indo-Aryan languages, such as Bengali, Gujarati, Marathi, and Punjabi. As a result of these borrowings, the Hindi sound system has a large number of consonants and vowels, as shown in Table 1.

Like many Indo-European languages, Hindi has parts of speeches such as noun and pronoun, verb, adjective, adverb, postposition, and conjunction. Hindi nouns have two genders, masculine and feminine. Though these categories are largely conventional and do not necessarily correspond to natural gender, there is some element of semantic consistency for many nouns and their gender, which depend on sex, size, shape, and degree of abstraction. However, many gender affiliations remain arbitrary.

The masculine gender nouns are associated with two different sets of stems and inflections and thus fall into two classes, *ā-stem*, such as g^hoṛā} ‘horse,’ and *non-ā-stem* nouns, such as nāg ‘snake.’ The feminine gender nouns, likewise, are associated with two different sets of stems and inflections and thus form two classes: feminine *i-stem* nouns such as nānī ‘maternal grandmother,’ and *non-i-stem* nouns, such as bahū ‘bride.’ Hindi nouns also occur in various cases represented by two inflected forms of nouns: a *direct form*, which represents the function of a subject or a direct object and an *oblique form* with its postposition, which represents all other syntactic functions or cases. It is actually the postposition that indicates the particular case. An oblique form of a noun with the postposition kō also functions as a direct or indirect object (accusative or dative). When a transitive verb occurs in the *perfective*, the special postposition nē occurs directly after the subject, causing it to appear in its oblique form. This subject with nē is said to be in *ergative* case. In *ergative* constructions the verb agrees in number and gender with its direct object if this occurs in its direct form, that is, without any postposition; otherwise the verb, occurring in the

Table 1

| Hindi consonants | Uvular | Glottal | Velar | Palatal | Retroflex | Alveolar | Dental | Labio-dental | Bi-labial |
|------------------|--------|---------|----------------|----------------|----------------|----------|----------------|--------------|----------------|
| Stops | | | | | | | | | |
| vls. | q | | k | | t | | t | | p |
| vls. asp. | | | k ^h | | t ^h | | t ^h | | p ^h |
| vd. | | | g | | ɖ | | d | | p ^h |
| vd. asp. | | | g ^h | | ɖ ^h | | d ^h | | b ^h |
| Fricatives | | | | | | | | | |
| vls. | | | χ | ʃ̌ | ʂ̌ | s | | f | |
| vd. | | h | ɦ | | | z | | | |
| Affricates | | | | | | | | | |
| vls. | | | | c | | | | | |
| vls.asp | | | | c ^h | | | | | |
| vd. | | | | j | | | | | |
| vd. asp. | | | | j ^h | | | | | |
| Nasals | | | | | | | | | |
| unasp. | | | ŋ | ɲ | ɳ | | n | ɳ | m |
| asp. | | | | | | | n ^h | | m ^h |
| Liquids | | | | | | | | | |
| unasp. | | | | | l | ɭ | l | | |
| asp. | | | | | ɭ ^h | r | l ^h | | |
| Semivowels | | | | y | | | | v | w |
| | | | | | | | | | |
| | Front | | | Back unrounded | | | Back rounded | | |
| | Oral | | Nasal | Oral | | Nasal | Oral | | Nasal |
| Hindi vowels | | | | | | | | | |
| high short | i | | ɨ | | | | | u | u |
| high long | ī | | ī̃ | | | | | ū | ū |
| mid short | e | | e̥ | | | | | o | o |
| mid long | ē | | ē̃ | | a | â | | ō | ō̃ |
| low short | | | | | ā | ā̃ | | | |
| low long | ḛ | | ḛ̃ | | | | | ṁ | ṁ̃ |

masculine singular form, agrees with neither the subject nor the object. The non-ergative types of sentences in Hindi are the usual nominative-accusative type, in which the verb agrees with the subject in number and gender.

Hindi verbs are associated with the categories of *tense* (present, past, and future), *aspect*, (imperfective, perfective, and continuous), and *mood* (indicative, subjunctive, and imperative). These categories encode grammatical aspects of meaning, some quite specific and some relatively vague or generic.

In Hindi, in addition to *derivation*, by which new words are formed from existing words through affixation, for example, *namak* 'salt,' *namkīn* 'salty,' new words are also formed from existing words by processes known as *compounding*, for example, *lāl pagaṛī* 'red turban: a policeman,' and *reduplication*. To express the ideas of *continuance distribution*, *variety*, *exclusion*, or *emphasis*, words are often completely or partially reduplicated in Hindi, for example, *bah-bah kar* 'floating floating down: repeatedly floating,' *g^har-g^har* 'house-house: each house,' *dēṣ-dēṣ* 'country-country: various countries,' *garam-garam* 'hot-hot: real hot,' and *badām-badām* 'almonds-almonds: only almonds.'

As it can be seen from the following examples, Hindi is an SOV (Subject-Object-Verb) language and sentences in Hindi show a threefold agreement involving the inflectional categories of number, gender, person, and case. These agreements are (a) between a modifier (adjective, adjective participle, and genitive attributive) and the noun it modifies, for example, *baṛā g^hoṛā* 'the big horse,' *baṛē g^hoṛē*

'the big horses,' *baṛī g^hoṛī* 'the big mare'; (b) between a predicative adjective and its subject, for example, (*g^hoṛā baṛā hæ* 'the horse is big,' *g^hoṛē baṛē hæ*) 'the horses are big,' *g^hoṛī baṛī hæ* 'the mare is big'; and (c) between a finite or main verb with its subject or object noun or pronoun, for example, *mē nē g^hoṛā xaridā* 'I-ergative horse-masculine bought-masculine: I bought a horse,' *mē nē g^hoṛī xaridī* 'I-ergative mare-feminine bought-feminine: I bought a mare,' *mē dōṛā* 'I-masculine ran masculine,' *tum dōṛē* 'you-masculine ran masculine,' *tum dōṛī* 'you-feminine anemone.'

Bibliography

- Beams J (1872–1879). *A comparative grammar of the modern Aryan languages of India. I, II, and III*. London: Trubner.
- Grierson G A (1903–1928). *Linguistic survey of India*. [Reprinted Delhi: Motilal Banarsidass.]
- Kellogg S H (1875, 1893). *A grammar of the Hindi language*. London: Routledge & Kegan Paul Ltd.
- McGregor R S (1977). *Outline of Hindi grammar*. Delhi: Oxford University Press.
- McGregor R S (1993). *The Oxford Hindi-English dictionary*. Delhi: Oxford University Press.
- Sharma A (1975). *A basic grammar of Hindi*. New Delhi: Central Hindi Directorate New Delhi.
- Shukla S (1981). *Bhojpuri grammar*. Washington, DC: Georgetown University Press.
- Shukla S (2000). *Hindi phonology*. Munich: Lincom Europa Academic Publishers.
- Shukla S (2001). *Hindi morphology*. Munich: Lincom Europa Academic Publishers.

Hindustani

H Dua, Central Institute of Indian Languages, Mysore, India

© 2006 Elsevier Ltd. All rights reserved.

Hindustani is a Central Indo-Aryan language based on Khari Boli (Khari Boli). Its origin, development, and function reflect the dynamics of the sociolinguistic contact situation from which it emerged as a colloquial speech. It is inextricably linked with the emergence and standardization of Urdu and Hindi. The linguistic relationship among Hindustani, Urdu, and Hindi highlights the theoretical and empirical problems of linguistic analysis and description. It also

reveals the politics of language conflict and identity in the complex sociopolitical and multilingual situation of India.

Origin and Development

Hindustani as a colloquial speech developed over almost seven centuries from 1100 to 1800. The Muslims conquered northern India from the 10th to the 13th centuries and settled down in the country, bringing with them their Persian language and culture. This mixing of cultures provided the contact situation for the emergence of Hindustani as a lingua franca. During this period, the literary language Apabhramśa seemed to be in a state of transition

from Middle Indo-Aryan to the New Indo-Aryan stage. Some elements of early Hindustani appear in compositions of the saints of Nath Panth. However, the distinct form of the lingua franca Hindustani appears in the writings of Amir Khusro (1253–1325), who called it Hindwi. Chatterji (1960) argued that the term ‘Hindustani’ came to be used at the close of the 17th century. J.J. Ketelaer is said to have written the first European grammar of Hindustani in Dutch in 1715.

During the early stage of language contact, Hindustani showed a great deal of mixing of dialects. In particular, the western dialect group of northern India – Braj Bhasha, Bangaru (Haryanvi), and to some extent eastern Punjabi – formed the basis of the vernacular Hindustani. The Perso-Arabic words were also used by traders, religious men, Indian Muslim nobles, and common people. It needs to be emphasized that the Khari Boli base of Hindustani was firmly established only in the 18th century. It is also important to note that no long and connected specimen of the language is available during the period 1200–1650 for reconstructing its continuous history, except *Bikat Kahani* by Afzal, which appeared 300 years after Khusro.

The next important phase in the development of Hindustani is seen in the Muslim Sultanates of the Deccan by the end of the 14th century. In the 17th century, Hindustani flourished as a literary language of the North Indian Muslims who settled in the Deccan. It is referred to as Dakhini Hindi or Urdu. Although it shows some elements of local languages, such as Kannada and Telugu, it clearly attests to the source dialects of Panjabi and Haryanvi. Furthermore, the literary works produced in Dakhini or Hindustani were written in the Perso-Arabic script, which “fixed the orientation of the language,” though Hindustani retained its indigenous character (Chatterji, 1960). The use of the Perso-Arabic script had serious implications for the development of Hindustani in the late 18th and 19th centuries.

The development of Hindustani in North India lagged behind its use as a literary language in the Deccan. One of the reasons for this was the recognition of Persian as an official and court language and its acceptance by the North Indian elites and court nobles. Second, Braj Bhasha flourished as a literary language there. It was recognized by the Mughul emperor Akbar, and his courtier Khan Khanan Rahim used it for his poetic compositions. Though after Khusro an early form of Hindustani can be found in the poetry of Kabir and other religious preachers and saints, it was not cultivated as a literary language. Furthermore, Hindustani was spoken among the nobles at Delhi and Agra and by Moghul emperors

from Akbar onward at home. However, it was not taken up seriously or written in the Perso-Arabic script in North India as in the Deccan.

It was only after Wali, a poet of Dakhini, arrived in Delhi that Hindustani began to develop as a literary language in the North. Wali used what is known as Rekhta/Hindi and showed that it was capable of great poetry. Rekhta means ‘scattered’ and implies that it had not yet so much been Persianized as happened later. It is known as the earliest form of Urdu-Hindustani poetical speech. Urdu as a language name occurs for the first time in 1776 in a couplet by the poet Mashafi (1750–1824). However, the use of Urdu referring to camp, court, or city (Zaban-e-Urdu or Zaban-e-Urdu-e-Shahi or Zaban-e-Urdu-e-Mualla) had been current since 1560.

After Wali, such stalwarts as Khan Arzu (1689–1756), Shah Hatim (1699–1781), and Mazhar Janejani (1700–1781) made conscious efforts to Persianize Hindustani and weed out the Braj Bhasha or indigenous elements from it. Thus, Urdu-Hindustani was in a developed state by the end of the 18th century. John Gilchrist produced the first grammar of Hindustani. He justified the usage of the term ‘Hindustani’ for the language, as well as for its speakers. Insha Allah Khan Insha’s *Darya-e-Latafat* (*The river of elegance*, 1807) presented an early linguistic study of the dialects of Delhi and Lucknow. There was also a tendency to identify Urdu-Hindustani largely as a Muslim language and Hindi/Hindwi as a language of the Hindus.

However, it is important to note two points about the development of Hindustani at the beginning of the 19th century. First, after the establishment of Fort William College by the British, prose began to be written in the emergent Khari Boli that formed the basis of Hindustani. Urdu-Hindustani and Hindi developed as two distinct styles of prose produced by the writers associated with Fort William College. Chatterji (1960: 211–212) rightly remarked that Hindustani “came out into the modern world as a vehicle of prose in its twin form, High Hindi and Urdu, about 1800.” Second, the process of identification of Urdu with Muslim and of Hindi with Hindus continued during the entire 19th century and the first half of the 20th century. The development and standardization of both Urdu and Hindi had sociopolitical implications with regard to Hindustani (see Urdu and Hindi).

Forms of Hindustani

It is quite clear that the evolution of Hindustani was spread over seven centuries from 1100–1800. Before it was firmly established on the Khari Boli spoken in

the surrounding region of Delhi, it was known by several names, such as Hindwi/Hindi, Dehalvi, and Rekhta, and was made up of several western dialects mixed together. The two forms of High Hindi and Urdu that emerged by 1800 may be described as a standardization of the grammar of the “Vernacular Hindustani” dialect (Chatterji, 1960: 169). The Perso-Arabic script and Perso-Arabic vocabulary of Urdu distinguish it from High Hindi, which uses Devanagari script and Sanskrit vocabulary.

The third form of Hindustani represents the basic Khari Boli and may be considered as Hindustani proper. It holds a balance in its vocabulary as it contains only those Persian and Sanskrit words that have been fully assimilated with the structure of its *tadbhav* or native words. According to Chatterji (1960), this form of Hindustani represents colloquial speech and can be terse as well as elaborate. It is simple in grammatical structure and precise in its sounds. In its spoken colloquial form, it is used for communication by a large number of speakers in India, Pakistan, and other parts of the world. It may therefore be considered, according to Chatterji (1960), as one of the great languages of the world.

Three other forms of Hindustani may be recognized, though they show a great deal of mixture from the local dialects and simplification of grammar. First, speakers of the Western Uttar Pradesh, Eastern Panjab, Haryana, and Rajasthan may speak what may be referred to as Vernacular Hindustani with their dialect accent or other features. They may be considered, as Kelkar (1968) maintained, to be ‘adherent’ speakers of Hindi and Urdu who easily ascend the scale of culture and education and accept them as super-posed languages. Second, it is possible to recognize what may be referred to as Bazaar Hindustani spoken by the masses in market situations across the country. This form may show a simplification of grammatical gender and mixing of local dialects, depending upon the region and the language contact situation. Finally, the Dakhini spoken in Karnataka, Andhra Pradesh, and other regions in the South may be considered to be a form of Hindustani. Though it has a distinctive grammatical structure, in vocabulary it shows affinity with Hindustani. It is spoken mainly at home and shows some local literary activity. However, the Dakhini speakers regard standard Urdu as the super-posed variety and have accepted it for all formal purposes.

Hindustani as a Symbol of Unity

The process of the identification of Urdu and Hindi with Muslims and Hindus, respectively, that started in the early 19th century reached its culmination in

the first quarter of the 20th century. The formation of voluntary language associations for these languages and the development of both Muslim and Hindu revivalism strengthened this identification. This congruence of linguistic and religious identities not only increased language conflict between Urdu and Hindi but also led to the expanded use of Hindustani. In the wake of the Indian independence movement, Gandhi saw the potential for the use of Urdu and Hindi to produce political conflict and so promoted Hindustani as a symbol of unity. In 1925, he persuaded the Indian National Congress to accept Hindustani as the official language for its proceedings. Under the influence of Gandhi, many national leaders emphasized the role of Hindustani not only for communal harmony between Muslims and Hindus but also for bringing about national unity. In 1937, Nehru recognized the potential of Hindustani to spread all over the country and declared that it should be officially recognized as an all-India language.

However, the acceptance and rapid spread of Hindustani at both regional and national levels failed to bring about any fundamental change in the position of the protagonists of Urdu and Hindi. The divergence between Urdu and Hindi languages, on the one hand, and the congruence of linguistic and religious identities, on the other, became so salient politically in the process of nationalism and nation formation that Hindustani failed as a symbol of unity. Das Gupta (1970:57) pointed out that the identification of national, linguistic, and religious solidarity was “more integral and pervasive” in the case of Muslims than with the Hindus.

After it became clear that India would be partitioned along religious lines, the question of Hindustani took a different turn in the fourth Constituent Assembly session in July, 1947. The persuasiveness of Hindustani as a national language had lost much of its appeal. The supporters of Hindi saw it as “a symbol of appeasement of the Muslim concern for Urdu” (Das Gupta, 1970: 130–131). They gave up their support for Hindustani and demanded that Hindi alone, written in the Devanagari script, be accepted as the official language of India. The acceptance of Hindi as the official language of India in 1948 gave the final blow to Hindustani as a symbol of unity. However, Hindi lost the overall support that Hindustani had gained at the national level in the wake of the independence movement. Recent debate on the failure of Hindustani as symbol of unity and a common language of both Urdu and Hindi speakers throws light on politics of nationalism, language engineering and acrimony between the two communities blaming one another for this (Rai, 2000; Hasnain and Rajyashree, 2004; Trivedi, 2004).

Problems of Linguistic Description

The problems of linguistic description of Hindustani are inextricably involved with those of Hindi and Urdu. Linguists, historiographers of Hindi and Urdu, and scholars of literature and textual criticism have mainly tried to grapple with the linguistic description of Hindi and Urdu, as is evident from the phonological and grammatical studies of those languages done during the last half-century. Kelkar (1968:1) has argued that contemporary standard Hindi-Urdu “consists of a gamut of integrated variation that need to be studied together – within a single frame work.” However, he concentrated mainly on standard Hindi-Urdu and considered Hindustani as “relegated to history” (Kelkar, 1968: 9), though he included it under the Hindi-Urdu continuum of styles and took care of regional color to a certain extent.

Though the linguistic description of Hindustani has not drawn the attention of linguists for historical reasons, it raises several theoretical and empirical issues, which are relevant for linguistic analysis of both Hindi-Urdu and Hindustani. First, in addition to the dictionary of Hindustani written by John Gilchrist at the end of the 18th century, several other dictionaries were published in the late 19th century. It would be relevant to explore the range of borrowed Perso-Arabic words included in the dictionaries and to find out how far they have been assimilated or have become current in present-day colloquial speech. Second, the works of several writers have been published in both Urdu and Hindi and have been claimed equally as Hindi-Urdu writers. Prem Chand occupies an important position in this respect. It would be worthwhile to explore the distinctive alternative use of Sanskrit or Persian words in Hindi and Urdu versions of his works and to study whether the common vocabulary comprises native or *tadbhava* or fully assimilated Sanskrit and Persian words. Third, a number of textbooks have been published for teaching standard Hindi and Urdu to foreigners. Although they reflect common core grammar and distinctive characteristics of the respective languages, it would be relevant to study the range of Sanskrit and Perso-Arabic words that form an integral component of these languages. Doing so would help determine how far these textbooks support the common base of colloquial Hindustani. Fourth, it would be necessary to explore to what extent the linguistic analysis of Hindi and Urdu is based on the spoken data. Only this type of analysis can show the extent to which they differ in the choice of Sanskrit and Persian words and to what

extent these words are common in both the spoken varieties and represent the colloquial Hindustani.

Finally, some studies show lexical differences between Hindi and Urdu and question the notion that they are two distinct languages. They raise significant issues related to the processes of convergence and divergence, the difficulty of drawing boundaries between Sanskrit and Perso-Arabic words assimilated in both Hindi and Urdu, and the implications of choice for comprehension. These issues can be explored only on the basis of a large corpus. A corpus of 3 million words is now available for both Hindi and Urdu at the Central Institute of Indian Languages in Mysore. On the basis of a comprehensive sample, it would be possible to explore in what kinds of genres/texts both Hindi and Urdu show a common base of colloquial Hindustani and how they differ from one another, on the one hand, and from Hindustani, on the other, in terms of what kinds of Sanskrit and Perso-Arabic words they use. In short, the research on the issues raised above can bring an understanding of the basic linguistic structure of Hindustani and the superimposed structure of Hindi and Urdu that is characteristic of both the spoken and written styles.

Bibliography

- Chatterji S K (1960). *Indo-Aryan and Hindi*. Calcutta: K. L. Mukhopadhyay.
- Das Gupta J (1970). *Language conflict and national development: group politics and national language policy in India*. Berkeley: University of California Press.
- Dua H R (1995). ‘Sociolinguistic processes in the standardization of Hindi-Urdu.’ In Hasnain I S (ed.) *Standardization and modernization: dynamics of language planning*. New Delhi: Bahari Publications.
- Faruqi S R (2001). *Early Urdu literary culture and history*. New Delhi: Oxford University Press.
- Hasnain S I & Rajyashree K S (2004). Hindustani as an Anxiety Between Hindi-Urdu Commitment. In Singh R (ed.) *The yearbook of South Asian languages and linguistics*. Berlin, New York: Mouton de Gruyter. 247–264.
- Kelkar A R (1968). *Studies in Hindi-Urdu: introduction and word phonology*. Poona: Deccan College.
- Mobbs M C (1981). ‘Two languages or one?’ *Journal of Multilingual and Multicultural Development* 2(3), 203–211.
- Rai Alok (2000). *Hindi nationalism*. Hyderabad, Delhi: Orient, Longman.
- Rai Amrit (1984). *A house divided: the origin and development of Hindi/Hindwi*. Delhi: Oxford University Press.
- Trivedi H (2004). The anxiety of Hindustani. In Singh R (ed.) *The yearbook of South Asian languages and linguistics*. Berlin, New York: Mouton de Gruyter. 267–273.

Hiri Motu

L Todd, University of Ulster, Coleraine, Northern Ireland

© 1994 Elsevier Ltd. All rights reserved.

Hiri Motu is a pidginized form of the Papuan language, Motu. It developed as a trade language between speakers of Austronesian and non-Austronesian languages. (*Hiri* is the Motu word for a trading expedition.) Hiri Motu almost certainly arose prior to European contact, but it has spread rapidly in the twentieth century, partly because of an increase in trading between Papuans and non-Papuans, and also partly because, under the name of *Police Motu*, it became the lingua franca of the multilingual police force, when Papua became an Australian Protectorate after World War I. It is an official language of Papua New Guinea, is used in politics and administration in the Papuan section of the country, and is spoken by about 200 000 people. It is regularly employed in the media, and has a standardized orthography, taken over from Motu, which was adopted in the nineteenth century as a church language.

Hiri Motu speakers use 5 vowel sounds, all of which can be combined, and 12 consonants /p, b, t, d, k, g, m, n, v, l/r, s, h/. All indigenous words end in a vowel, *dubu* 'church,' *turana* 'friend,' and the consonants /l/ and /r/ are interchangeable for the majority of speakers.

The similarity between Hiri Motu and other Pacific languages may be illustrated by comparing a few items of Hiri Motu's vocabulary with their cognates in Hawaiian (1):

| Hawaiian | Hiri Motu | English | (1) |
|---------------|---------------|----------|-----|
| <i>kalo</i> | <i>taro</i> | 'taro' | |
| <i>lau</i> | <i>raurau</i> | 'leaf' | |
| <i>wahine</i> | <i>hahine</i> | 'woman.' | |

Words which are borrowed from English are restructured to suit Hiri Motu's phonology, *besini* < *basin*, *botolo* < *bottle*, *sopu* < *soap*, *tosi* < *torch*.

For Papuans, sentence structure usually follows the pattern for Motu. It is OSV when the subject is a pronoun (2):

Turana ia itaia. 'friend he see' (2)
= 'He sees a friend.'

or a SOP when the subject is a noun phrase (3):

Kuku ese aniani ia nadu. 'cook + (3)
particular food he boil'
= 'The cook boils the food.'

Adjectives follow nouns (4):

hanua ta *hanua toi* (4)
'one village' 'three villages.'

Time is either implied from the context or indicated by means of auxiliaries (5):

Au do lau utua. 'tree future I cut' (5)
= 'I'll cut the tree.'
Au lau utua vadaeni. 'tree I cut past'
= 'I've cut the tree.'

The omnipurpose postposition *dekenai* follows the noun (6):

dala dekenai 'on the road' (6)
dubu dekenai 'to church'
ruma dekenai 'in the house'

Europeans and New Guinea speakers tend to impose the grammatical patterns of their mother tongues or those of Tok Pisin on Hiri Motu. This is particularly true with regard to word order, which is SPO for many.

The government of Papua New Guinea tends to promote Tok Pisin and Hiri Motu equally, partly to avoid ethnic tensions. Before unification and independence, which occurred on September 16, 1975, Tok Pisin was most frequently used by New Guineans and Hiri Motu by Papuans.

Bibliography

- Dutton T E & Voortoeve C L (1974). *Beginning Hiri Motu*. Canberra: Pacific Linguistics, Australian National University.
Todd L (1984). *Modern Englishes: Pidgins and Creoles*. Oxford: Blackwell.
Todd L (1989). *Pidgins and Creoles*, 2nd edn. London: Routledge.

Hittite

J G Macqueen

© 2006 Elsevier Ltd. All rights reserved.

Hittite is the name now given – its users probably called it Nesite – to the Indo-European language employed in the north-central area of Anatolia (modern Turkey) during much of the second millennium B.C. Large numbers of texts have been excavated, principally at the site of Boğazköy (Boğazkale), the ancient Hattusas, capital of the Hittite kingdom. As evidence for the language comes entirely from documentary sources, the number of speakers cannot be estimated; but it is probable that it was also the spoken vernacular of the area, at least from ca. 1700 to ca. 1300 B.C. During that period, it shows normal signs of linguistic change, but there is evidence, in the absence of continued change and in the increasing presence in the texts of forms which are Luwian in origin, to suggest that after ca. 1300 B.C. it had become a ‘dead’ language, its use confined to the Hittite chancellery, and that its more southerly relative replaced it as a vernacular.

Phonology

Hittite is written in a local variety of the Mesopotamian cuneiform script. This script, being basically unsuited to the language, makes full understanding of many phonetic features (e.g., vowel length; voiced versus voiceless stops) difficult. The vowel system shows *a* (which reflects both I-E *a* and I-E *o*), *e* and *i* (which are distinguished in earlier texts, but may well in later ones represent a single phoneme), and *u*. There are four orders of stops (labial, dental, velar, labio-velar), and lack of voice is often indicated by gemination in spelling, although the converse is by no means always the case. Hittite is unique among Indo-European languages in preserving the continuants known as laryngeals. The number and precise nature of these is still the subject of debate; but both voiced and voiceless varieties can be detected, and in some cases they can be seen to give *a*- or *o*-coloration to an original *e* vowel.

Morphology

Nominal morphology is characterized by the loss of both the feminine gender and the dual number. Earlier texts show a full range of case forms in the singular, although in later texts dative and locative have merged. In the plural, the range of case forms is much reduced. In noun formation, a striking

characteristic is the preservation in productive use of heteroclitic *r/n* stem neuters and of action nouns in *-sar*, *-tar*, and *-warl-mar*.

The verbal system shows two tenses (present and preterite), two moods (indicative and imperative), and two voices (active and medio-passive), together with two infinitives, a supine, a verbal noun, and a participle. The present indicative of the medio-passive is often marked by a suffixed *-ri* which links it to the medio-passive in Celtic, Italic, and Tocharian. There are also two conjugations (the *mi*- conjugation and the *hi*- conjugation), the first of which corresponds to the I-E present while the second is perfect in origin with the addition of a present marker.

Syntax

Syntax is on the whole simple and straightforward, and corresponds in general to that of the archaic forms of other I-E languages. Characteristic of Hittite is a liking for ‘chains’ of particles and enclitic pronouns placed at the beginning of a sentence or clause. Another interesting feature is the ‘quasi-ergative’ construction, in which a neuter subject with a transitive verb is not permitted, and so an original ablative case is reinterpreted as a nominative.

Example

A Hittite sentence showing some of the above features is (in syllabic transcription)

- (1) *an-da-ma-za pa-aḫ-ḫu-u-e-na-aš-ša u-da-ni-i*
me-ik-ki na-aḫ-ḫa-an-te-eš e-eš-tin,

to be read

- (2) *anda-ma-tsa pahhwenas-a uddani mekki*
nabhantes esten

and translated

- (3) ‘moreover, on the subject of fire also be greatly fearful,’

In this sentence, *anda* acts as an adverb of transition, but is ultimately linked to the form seen in Greek *éndon*, Old Latin *endo*; *ma* is a connective particle; *tsa* is a reflexive particle; *pahhwenas* is the genitive of *pahhur*, an *r/n* stem noun cognate with Greek *pûr* and showing the presence of an original laryngeal; *-a* is an emphasizing particle; *uddani* is the dative-locative of another *r/n* stem noun with a basic meaning of ‘word’ and a possible ultimate connection with an I-E verb of saying (cf. Old Welsh *dy-wedut*); *mekki* is an adverbial neuter singular of an adverb

meaning 'much, many' (cf. Skt *mah-*, Gk *mégas*); *nabhantes* is the participle (cf. Latin *amans*, *amantis*) of a verb perhaps cognate with Old Irish *nar*, 'timid'; and *esten* is the 2nd pl imperative of the I-E verb for 'to be.'

Records in the Hittite language cease with the collapse of Hittite power ca. 1180 B.C. There are signs that the language of classical Lydia is a later relative; but the precise relationship is obscure.

Bibliography

- Friedrich J (1960). *Hethitisches Elementarbuch I*. Heidelberg: C. Winter.
 Guterbock H J & Hoffner H A (1980 onwards). *The Chicago Hittite Dictionary*. Chicago, IL: Oriental Institute.
 Macqueen J G (1986). *The Hittites* (2nd edn.). London: Thames and Hudson.
 Puhvel J (1984–1990). *Hittite Etymological dictionary*. Berlin: Mouton de Gruyter.

Hmong-Mien Languages

M Ratliff, Wayne State University, Detroit, MI, USA

© 1994 Elsevier Ltd. All rights reserved.

The Hmong-Mien (= Miao-Yao) language family comprises some 30–40 languages spoken primarily in southwestern China, but also in northern Vietnam, Laos, and Thailand. Three main branches have been identified: Hmongic (= Miao), an internally diverse subfamily, including among others the languages Hmong, Bunu, Hmu, and Qo Xiong; Mienic (= Yao), a smaller and less diverse subfamily, including Mien and Mun; and Ho Nte (= She), consisting of the Ho Nte language alone. Further research on lesser-known members of the family may lead to the identification of more branches. The family designation *Miao-Yao* is of Chinese origin, and represents an ethnic classification rather than a purely linguistic one. Primarily for this reason, many Western scholars have adopted the name *Hmong-Mien* to refer to this language family. Genetic relationships to Sino-Tibetan, Austro-Tai, and Austric have been proposed. Due to typological similarities shared by member languages of the four main families represented in Southeast Asia (Sino-Tibetan, Hmong-Mien, Tai-Kadai, and Mon-Khmer) and contact-induced borrowings, however, it is difficult to establish the distant relations of the family with confidence.

The 1982 census in China reported 4.5 million speakers of Hmongic languages and 750 000 speakers of Mienic languages (only approximately 1000 speakers of Ho Nte live in Guangdong province near Hong Kong). They inhabit Guizhou, Guangxi, Hunan, and Yunnan provinces, and have a lesser presence in Sichuan, Guangdong, Hubei, and Jiangxi provinces and the island of Hainan. From the early nineteenth century through the early twentieth century, speakers of the Hmong, Mien, and Mun languages

moved in waves into northern Southeast Asia under pressure from the expanding Han population. The ratio of Hmong-Mien speakers in China to those in northern Southeast Asia is now approximately 5:1. Finally, there was further displacement of tens of thousands of Southeast Asian Hmong and Mien following the end of the Indochinese war in the mid-1970s, primarily to the USA, France/French Guiana, and Australia.

Speakers of Hmongic and Mienic languages have long been dominated by speakers of Chinese. There are consequently typological similarities between members of the two families. Hmong-Mien languages have monosyllabic morphemes, which occur freely or in transparent compounds, and very little affixal morphology. They are characterized by the presence of numeral classifiers, serial verb constructions, zero anaphora, expressives (ideophones), and sentence particles expressing a variety of pragmatic functions. All the languages are tone languages, some with world-record complexity: Shidongkou Hmu (= Black Miao) has been reported to have five level tones, for example, and Longmo and Zongdi Hmong each have 12 tonal contrasts. Hmongic languages are characterized by extremely rich initial consonantism (including retroflex and uvular places of articulation; prenasalized, aspirated, and glottalized stops; voiceless sonorants) and impoverished final consonantism, whereas Mienic languages are characterized by up to six consonant contrasts syllable-finally (-m, -n, -ŋ, -p, -t, -k), a rich system for the area, and correspondingly fewer initial contrasts.

Bibliography

- Strecker D (1987). The Hmong-Mien languages. *Linguistics of the Tibeto-Burman Area* 10(2), 1–11.
 Wurm S A (1987). *Language Atlas of China*. Hong Kong: Longman.

Hokan Languages

T Kaufman, University of Pittsburgh,
Forest Hills, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

History of Scholarship

The Hokan Hypothesis

The Hokan hypothesis (the hypothesis that the Hokan languages are a genetic group descending from a common protolanguage) was the result of a taxonomically-motivated attempt in 1912–1913 by Roland P. Dixon and Alfred Louis Kroeber to deal with the very large number of apparently distinct genetic groupings of languages (according to John Wesley Powell's 1890 classification) known for aboriginal California – large in comparison with other parts of North America. Dixon and Kroeber, who had moderate amounts of (phonologically not very accurate) data from, and moderate amounts of familiarity with, a large number of California languages, felt that by looking at all the languages panoramically and grouping them by overall shared lexical similarities, a set of groupings could be achieved that might as well be viewed as genetic. This exercise yielded two primary sets of languages that Dixon and Kroeber labeled Penutian and Hokan, as well as some ungrouped languages. Penutian included Yokutsan (Yokuts), Miwokan, Costanoan, Wintuan (Wintu), and Maiduan; Hokan included Karuk (Karok), Shastan, Achumawi-Atsugewi (Achumawi), Pomoan (Pomo, Southeastern), Yanan (Yana), Esalen (Esselen), Salina (Salinan), Washu (Washo), and Yuman (Dixon and Kroeber, 1913a, 1913b).

Additions to the Hokan 'Core'

Later Kroeber added Seri and Chontal to Hokan (Kroeber, 1915). He (?Sapir) misguidedly tried to add Chumashan (Chumash) to Hokan. Edward Sapir, the most able American linguist between 1905 and 1925, who had done doctoral research on Takelma, added Takelma, Klamath (Klamath-Madoc), Sahaptian (Sahaptin), Alsea, Kusan (Coos), Chinookan (Chinook), Tsimshian, and Sayusla (Siuslaw) to Penutian. He added Pajalat, Yeme*, and Yue* to Hokan, but his efforts at adding Karankawe were misguided. He misguidedly tried to add Sutiaba to Hokan. In 1953, Greenberg and Swadesh added Tol to Hokan.

Hokan as a Superstock or Phylum

When originally proposed, the Hokan and Penutian hypotheses were not directly comparable to hypotheses of the order of Algonkian or Yuta-Nawan, because the numbers of resemblant forms that could

be deployed and the sound correspondences that could be discerned were so few that any true genetic relationship that lay behind the phenomena was necessarily very remote. No certainty could be felt that such a hypothesis would ever be demonstrated to be correct through reconstruction. The terms 'family' and 'stock' had been used by comparative linguists since before 1800, but by 1900, most linguists felt these terms should be reserved for genetic groupings that were not in doubt. Consequently, the terms 'phylum' and 'superstock' began to be used between 1930 and 1960 to refer to such hypotheses as Hokan and Penutian. Proposed genetic groupings were often preceded by the prepound 'macro-,' thus Macro-Hokan, Macro-Penutian, and Macro-Chibchan. Doubters of the validity of the Hokan hypothesis may label it Macro-Hokan; those who believe in the hypothesis will speak of the Hokan stock.

Further Comparative Studies

Except for a small number of proposed reconstructions by Sapir, none of the above studies did more than assemble sets of resemblant forms with comparable (often identical) glosses or functions.

Since 1950, renewed efforts at establishing Hokan or parts of it have been devoted by linguists working on documenting these languages, especially as dissertations by graduate students at University of California at Berkeley (UC Berkeley). The Hokan comparative studies have been spin-offs of their documentation work. Even now, however, except for the reconstruction of Pomoan and Yuman, Hokan comparative studies at most surpass those of the period 1900–1940 by attempting to find sound correspondences among the resemblant forms compared. The other main achievement is that the data that are compared are for the most part phonologically accurate.

The Hokan stock is made up of the following (reasonably) well-documented languages and families: (1) Pomoan family; (2) Chimariko language; (3) Yanan small family; (4) Karuk language; (5) Shastan (Shasta) language; (6) Achuan family; (7) Washu language; (8) Salina language; (9) Yuman family; (10) Seri language; (11) Chontal language area; (12) Tol (= Jicaque) small family.

Some other poorly-documented languages are probably Hokan, but because of lack of data they cannot serve as the basis of reconstruction: Esalen, Kochimi* (Cochimí), Pajalat, Yeme* = Komekrudo, maybe Yue* = Kotoname.

Others often thought of as Hokan are probably not Hokan: Chumashan, Tonkawa, Karankawe. In any case, the first two, which are reasonably well

documented, have not successfully been shown to be Hokan, in spite of Sapir's efforts.

Contra Sapir (1925) (aped in Greenberg, 1983) Tlapaneko-Sutiaba is Oto-Mangean, not Hokan. Whether Hokan and Oto-Mangean are related remains an open question. In Kaufman (1990), I suggest that they are indeed related.

To date, the most notable contributions to Hokan typology and comparative Hokan grammar have been made by Sapir, Jacobsen, Langdon, Gursky, Grey, and Oswalt.

Kaufman has a fair amount of evidence to suggest a North : South division, North being 1–9, and South being 10–17.

The Hokan hypothesis is widely known; it may be accepted by nonspecialists in American Indian languages, and is accepted (with a specific list of languages shorter than the list given in this article) by many specialists in languages of Oregon, California, and Meso-America. Specialists in American Indian languages with no deep familiarity with Hokan languages are generally skeptical of the Hokan hypothesis. This skepticism is largely because, although detailed comparative work leading to reconstruction has been carried out for some of the parts of the Hokan grouping (Yuman, Pomoan), comparative work at the level of the whole stock has not yet led to fully elaborated reconstruction of either phonology or grammar.

The Hokan languages are known from three separate areas: Alta California and Baja California; Southern Texas and Coahuila; Southeastern Oaxaca and Northern Honduras. Many languages of northwestern and northeastern Mexico have disappeared with essentially no documentation: some of these may have been Hokan. From the current distribution, we could imagine a Hokan homeland or primary geographical concentration in southern California with extensions to northern California and the southern plains (skipping over what?). The Hokan languages of Meso-America would have to represent migration, as might also those of the southern plains.

The time depth of Hokan is probably quite great, perhaps 8000 years, and the population movements that need to be postulated would possibly not have been associated with distinctive archeological traditions.

Evidence assembled by Kaufman suggests that Hokan and Oto-Mangean are genetically related. If so, proto-Oto-Mangean (ca. 6500 mc) would have developed from a Hokan-like language with some of the following changes:

1. syllable-final obstruents would have dropped or become laryngeals
2. surviving features of the changed consonants would have yielded a three-way tonal contrast

3. a thorough-going shift on the phrase and clause level from right-headedness (OV) to left-headedness (VO) would have affected all morpheme arrangements above the level of derivational morphology (which involved only full words and clitics).

The Hokan Languages Classified

In the following sections, % means that the language is extinct.

Northern Hokan

Sonoma

1. Pomo family [pPom] (total number of speakers < 200)
 - Western Pomo language area [WPom]
 - SouthWestern (Kashaya) Pomo emergent lg [SWPom] EL: ca. 50
 - Southern Pomo emergent lg [SPom] EL: < 40
 - Central Pomo (Yokaya & Boya) emergent lg [CPom] EL: < 40
 - %Northern Pomo lg [NPom]
 - NorthEastern (Salt) Pomo lg [NEPom] EL: 1
 - %Eastern Pomo lg [EPom]
 - SouthEastern (Sulphur Banks) Pomo lg [SEPom] EL: < 10

Northern California

2. %Chimariko [ch'imari*ko] language [Chi]
3. %Yana language area [pYan]
 - Yana emergent lg [Yan]
 - Northern Yana dial [NYan]
 - Central Yana dial [CYan]
 - Southern Yana dial [SYan]
 - Yahi emergent lg [Yah]
4. Karuk [karu*k] language [Kar] EL: 126
5. %Shastan family [pSha]
 - Shasta lg [Sha]
 - New River Shasta lg [NRSha]
 - Okwanchu lg [Okw]
 - Konomihu lg [Kon]
6. Achu family [pAch] (total number of speakers < 100)
 - Achumawi (Pit River) lg [Ach] EL: 81
 - Atsugewi (Hat Creek) lg [Ats] EL: 4
 - Atsuge (Hat Creek) dial [Ats-HC]
 - Apwaruge (Dixie Valley) dial [Ats-DV]

Great Basin

7. Washu language [Wsh] EL: < 10

California Coast

8. %Esalen language [Esa]
 %Salina (Enalen) language [Sal]
 Miguelenyo dial [Sal-M]
 Antonienyo dial [Sal-A]

Southern Hokan**Southwest**

9. Yuman-Kochimi* family [pYK] (total number of speakers ca. 2500)
 Yuman division [pYum]
 Pai language area [Pai]
 Paipai emergent lg [Pp] EL: 300 BAJA
 Havasupai-Hwalapai emergent lg
 Havasupai dial [Hav] EL: 404
 Hwalapai dial [Hua] EL: 440
 Yavapai emergent lg [Yav] EL: 163
 River language area [Riv]
 Mohave emergent lg [Moh] EL: 234
 Maricopa-Yuma emergent lg
 Maricopa dialect [Mar] EL: 181
 Yuma dialect [Yum] EL: 343
 Dieguenyo language area [Die] EL: 97
 Mesa Grande ('Iipay) emergent lg [MG]
 Campo (Kumeyaay) emergent lg [Cam]
 La Huerta (Tiipay) emergent lg [Hue]
 Cocopa lg [Coc] EL: 321
 Kiliwa lg [Kil] EL: 24-32 BAJA
 %Kochimi* language [Cch]
10. Seri language [Ser] EL: <215 BAJA

Coahuila

- %Pajalat (Coahuilteco) language [Paj]
 %Yeme*an (Comecrudoan) family [pYem]
 Yeme* (Comecrudo) lg [Yem]
 Garza lg [Gar]
 Mamulique lg [Mam]
 %Yue* (Cotoname) language [Yue]

Oaxaca

11. Chontal (Tequistlatecan) family [pCho]
 Huamelulteco = Lowland Chontal lg [LCho]
 EL: 1 k
 Tequistlateco = Highland Chontal lg [HCho]
 EL: 3.6 k

Honduras

12. Tol (Jicaque) family [pTol]
 Eastern Tol lg [ETol] EL: 350
 %Western Tol lg [WTol]

Descriptive Work on Hokan Languages

From 1900 to 1950, the documentation of Hokan languages that stands the test of time includes Edward Sapir's documentation of Yanan and John Peabody Harrington's documentation of Karuk, Chimariko, and Salina. Just before 1950, Abraham Halpern documented Yuman.

Since 1950, Hokan languages of Meso-America have been documented first by members of the Summer Institute of Linguistics and later by academic linguists from the United States.

Also since 1950, Hokan languages of Alta California (and Washu in Nevada) and Baja California have been the object of descriptive study leading to Ph.D. dissertations by students of the University of California at Berkeley. Since 1970, several linguists trained by Berkeley Ph.D.s have documented Hokan languages of Alta and Baja California.

Characteristics of Hokan Languages**Phonology**

Phonemic Contrasts in Hokan Languages To get an idea of what the broadly general traits of Hokan phonological systems are, I present below those phonological contrasts that are common or predominant in these languages.

1. /C'/ vs. /C/: pPom, Chi, Yan, Sha, Ach, Wsh, Sal, Paj, pCho, pTol no /C'/: Kar, pYum, Ser, [Yem unclear]
2. /C^h/ vs. /C/: pPom, Chi, Yan, pAch, Wsh, pTol no /C^h/: Kar, Sha, Sal, pYum, Ser, Paj, pCho, [Com unclear]
3. /t/ vs. /t/: pPom, Chi, Sal, [pYum] no /t/: the rest
4. /č/ vs. /ç/: Chi, Sha, Sal, Paj no contrast: the rest
5. /k^y/ vs. /k/ or /q/: pPom, pYum no contrast: the rest
6. /q/ vs. /k/: [pPom], Chi, Ach, pYum no contrast: the rest
7. /k^w/ vs. /k/: pYum, Paj, Com, pCho no /k^w/: the rest
8. /f/ vs. /p/: Kar, Ser, pCho no /f/: the rest
9. /š/ or /ʃ/ vs. /s/: pPom, Chi, [Kar], Ach, Wsh, Sal, pYum, Ser, Paj no contrast: Yan, Sha, Ats, Yem, pCho, pTol
10. /x^w/ vs. /x/: pYum, Paj, Yem, pCho no /x^w/: the rest
11. /h/ vs. /x/: [pPom], Chi, [Yan], Sha, Kar, Ach, [Sal] no contrast: the rest
12. /r/ vs. /l/: Yan, pAch, pYum no contrast: the rest
13. /e/ vs. /i/: pPom, Chi, Yan, Sha, Ach, Wsh, Ser, Paj, Yem, pCho, pTol no contrast: Kar, Ats, Sal, pYum

14. /o/ vs. /u/: pPom, Chi, Yan, Ach, Wsh, Paj, Yem, pCho, pTol no contrast: Kar, Sha, Ats, Sal, pYum, Ser
15. /ɨ/ vs. /i/ or /u/: Wsh, pTol no /ɨ/: the rest
16. vowel length: pPom, [Chi], Yan, Sha, pAch, Wsh, Sal, pYum, Ser, Paj no contrast: ?pCho, pTol
17. stress or pitch: pPom, Kar, Sha, pAch, Wsh, Sal, pYum, Ser, pCho no contrast: Chi, Yan, Yem, pTol.

A sort of common core that is predominant (though not universal) in the Hokan languages is shown in (Table 1).

Just in terms of the phonemic contrasts commonly found in Hokan languages, the set of phonemic contrasts shown in Table 2 is the maximum that is supported typologically (i.e., found in at least three branches).

Syllable Structure

A syllable may begin with a consonant or not; it may end with a consonant or not. Some languages tolerate syllable onsets of the shape /Cx/ or /CY/ (where Y is a semivowel); many languages tolerate syllable codas of the shape /YC/.

I postulate the following basic phonological structure for a proto-Hokan lexical item of one to three syllables that is not a compound (\$ is syllable boundary):

Table 1 Basic consonants and vowels in Hokan languages

| Consonants | Vowels |
|--------------------|------------|
| <i>p t c k</i> | <i>i u</i> |
| <i>p' t' c' k'</i> | <i>e o</i> |
| <i>s š x h</i> | <i>a</i> |
| <i>m n</i> | |
| <i>l</i> | length /:/ |
| <i>w y</i> | stress /*/ |

Note: only Kar, Yum, and Ser lack glottalized obstruents; only Cho and Tol lack vowel length; only Chi, Yan, Yem, and Tol lack contrastive stress.

Table 2 Phonemic contrasts in Hokan languages

| Consonants | Vowels |
|-------------------------------------|------------|
| <i>p t t̚ /c̚ /č k kʷ q</i> | <i>i u</i> |
| <i>pʰ tʰ t̚ʰ /c̚ʰ /čʰ kʰ kʷʰ qʰ</i> | <i>e o</i> |
| <i>p' t' t̚' /c̚' /č' k' kʷ' q'</i> | <i>a</i> |
| <i>f s š x xʷ h</i> | |
| <i>m n</i> | length /:/ |
| <i>l</i> | stress /*/ |
| <i>r</i> | |
| <i>y w</i> | |

- (1) \$([C(x)]V[H])\$C(x/w)V(H) (Y) (C)\$(+CV)

Grammar

Hokan morphology is typically OV, while several Hokan groups currently show VO syntax.

Grammatical traits that will be discussed are as follows: alignment and person markers, nouns, verbs, adjectives, interrogatives, and quantifiers; examination of these traits is followed by a discussion of word order.

Alignment and Person Markers Most Hokan languages (e.g., Yan, Yum, Ser) have accusative case-marking; some (e.g., Chi, C&E Pom, Cho) have active case-marking. While Ergative languages often have completely different sets of person markers for ergative versus absolutive case (Mayan, Philippine languages), Accusative (Yuta-Nawan [Uto-Aztecan] family, Sapotekan [Zapotecan] family) and Active (Siuan [Siouan] family, Masatekan [Popolocan] family) languages often show related or even identical markers for the case categories that encode Agent and Patient.

In Yanan, an Obj-Subj suffixed person-marking combination follows TAM markers on the verb. In Chontal, an Active person marker precedes the verb stem and a Neutral (= Stative) person marker follows all other verbal inflexional suffixes. In Yuman and Seri, a prefixed Obj-Subj combination precedes the verb. In Chimariko, Agent or Patient is prefixed to the verb (the category that is marked is chosen according to a person hierarchy). Since in Pom and Sal, verbs are not person-marked for subject and object agreement, it seems likely that proto-Hokan had no such marking. The Chontal order reflects its SIVO pattern (VO is favored by Meso-American languages: I means indirect object), and the Seri, Yuman, and Chimariko patterns reflect their current SV and OV word orders, which are probably the proto-Hokan orders as well, although the pattern with full NP arguments is specifically SOV, not OVS. The Yanan, Yuman, and Seri data suggest that in early Hokan there may have existed an O-V (O-S?) clitic combination for person markers. The Yanan order would then reflect the verb-first syntax of Yanan. If pHokan was an Active rather than an Accusative language, the alignment categories of the pronominal clitic combination would be Neutral-Active. See below for possessor marking on nouns.

Nouns The noun stem is made up of a root optionally followed by a nominalizer ('infinitive/gerund,'

‘agentive’), or a ‘first-order nominal suffix’ (Pom, Chi, Wsh, Esa, Yum: not productive) and a ‘second-order nominal suffix’ (‘diminutive,’ ‘female’). In Pomo and Chimariko, a noun root may be preceded by another noun root to form a noun-noun compound with the first noun modifying the second; Salina and Yuma have no such compounding, and the status of such compounds in proto-Hokan is still in doubt.

- (2) | (NOMLZ) |
 (ROOT) { ROOT } (NOM₂)
 | (NOM₁) |

The noun word consists of a noun stem plus up to two preposed optional grammatical markers, a ‘possession state prefix’ (‘absolutive of intimately possessed noun’/‘substance or mass noun prefix’ [Yum, Ser, Cho, Tol], ‘body-part prefix’/‘possessed state of intimately possessed noun’ [Pom, Chi, Yum, Ser, Cho], ‘indefinite third person (+/– possessive)’/‘absolutive noun prefix’ [Pom, Kar, Sha, Yan, Wsh, Sal, Yum, Ser]), and a ‘proclitic classifier’ (‘proclitic count noun article,’ ‘proclitic mass/plural noun article,’ ‘absolutive noun prefix’), one optional ‘possession state suffix’ (‘absolutive’) and one obligatory ‘case suffix.’ The classifier has become a prefix in many languages.

The cases are both locative and relational, but the only relational cases that have etymologies are limited to Southern Hokan and encode switch-reference (same subject vs. different subject).

- (3) (CLASS) (POSS) NOUNSTEM (POSS) CASE

Locative cases suffixes encode such functions as ‘by means of,’ ‘from,’ ‘at,’ ‘in,’ and so on.

Verbs The verb stem is made up of a root optionally preceded by a stativizer (‘adjective-like intransitive’), a causativizer, or an incorporated instrumental pre-pound (Pom, Chi, Yan [‘primary verbs’], ?Sha, *Kar, Ach, Ats, Wsh, Yum, Ser, *Cho) and optionally followed by a ‘first-order verbalizer’ (‘to do X [X = noun, numeral]) and an incorporated ‘directional postpound’ (Pom, Chi, Yan, Kar, Sha, Ats, Wsh, Yum, Cho).

Both Northern and Southern languages have instrumental prepounds. Only Salina seems definitely to lack them. Instrumental prepounds encode such meanings as ‘with the mouth,’ ‘by speaking,’ ‘by biting/chewing,’ ‘by blowing,’ ‘with the foot,’ ‘with the hand.’ They are seemingly recruited mostly from noun and verb roots.

Both Northern and Southern languages have directional/locative postpounds. Again, only Salina seems

definitely to lack them. Directional/locative postpounds encode such meanings as ‘down,’ ‘up,’ ‘in,’ ‘out,’ ‘away,’ ‘thither,’ ‘hither,’ ‘here and there.’

It is not clear whether there is widespread simple noun incorporation or verb root compounding. No incorporation is found in ?Pom, Sal, or Yum. V-N incorporation is found in Yan, but is probably an innovation reflecting its VO syntax. Langdon (1988) shows that compound verb stems whose first member is not necessarily instrumental occur in Yana, Shasta, Atsugewi, and Washu.

- (4) | (STAT) |
 { (CAUS) } ROOT (VRBLZ₁) (DIR)
 | (INSTR) |

The verb word (or verb complex) consists of a verb stem plus up to three preposed optional grammatical markers and up to five postposed grammatical markers, of which one is obligatory. The preposed markers are [–1] pluralizer prefix (Yum, Paj, Pom, Yan, Ats), [–2] future proclitic (Yan, Kar, Wsh, Sal, Ser, Cho, Paj, Tol), [–3] temporal subordinator proclitic (‘when, while, after’: Pom, Yum, Sal). The postposed markers are [+1] passivizer suffix (Sal, Yum, Ser) or SHIFTER (infinitive, agentive), [+2] andative suffix (‘to go and verb’: Northern Hokan only), [+3] obligatory TAM₁ suffix (e.g., imperative, present, future/optative, past/completive, remote past, desiderative), [+4] TAM₂ suffix (e.g., conditional), [+5] TAM₃ enclitic (e.g., customary/habitual). Verb-TAM order is found generally in Hokan: Pom, Chi, Yan, Kar, Wsh, Sal, Yum, Cho.

- (5) (TEMP SUBORD) (FUT) (PL) VERBSTEM
 (PASS) (ANDAT) TAM₁ (TAM₂) (TAM₃)

Adjectives Some adjective-like words act like nouns and some act like verbs. While it is possible that adjectives as a class originally had no independent existence, it is equally possible that there were two kinds of adjectives that were neither nouns nor verbs, as there are in Nahua (Yora), Mayan, and Bantu languages, to name just three cases.

Interrogatives In Amerindian languages and elsewhere, interrogative words are often encoded by lexical items that also have generic reference, such that *who?* = ‘person,’ *what?* = ‘thing,’ *where?* = ‘place,’ *how?* = ‘manner,’ and so on. This is also true of the interrogative words in Hokan languages.

Quantifiers Structurally, the numerical systems of Hokan languages do not reflect a widespread practice among their speakers of calculating or counting large

numbers of things. There are three widespread etyma that mean 'one' and/or 'only, alone.' There are three that mean 'three', and two that mean 'two' (The Oto-Mangean stock also shows multiple etyma for most of the low numerical values, and little evidence for numbers above five). Etyma with values above three are found only in Northern Hokan. The various words for 'two' in the Hokan languages do not all reflect a single unitary proto-Hokan etymon, and the invented word *Hok[an]* does not directly represent any of them, although such was Dixon and Kroeber's intention.

Word Order The following remarks are based on a structural survey of certain languages only: Pomo (p.c. Oswalt, McLendon), Chimariko (TK), Yana (p.c. Hinton), Salina (p.c. Turner), and Yuma (p.c. Hinton, Langdon). Before much more can be done in this area, syntactic descriptions of Karuk (Bright), Shasta (Silver), Washu (Jacobsen), and Seri (Marlett) will have to be consulted, and descriptions of Achumawi, Atsugewi, Chontal, and Tol will have to become available.

Sentence-level Constituents On the level of the sentence, SOV word order is attested from Pomoan, Chimariko, Yuman, and Seri, and proto-Hokan probably had this order as well. Yana has VSO, and Salina has VOS.

Noun Phrase Within the NP, the modifying adjective probably originally followed the noun it modified. Pomo, Chimariko, Yuman, and Seri all attest this, though Salina has AN order. As is well-known by now, NA order is neutral with respect to OV or VO constituent order, and not unharmonious with OV order.

There are two kinds of possessive constructions: one where the possessor [G] is an NP and one where it is a first or second person pronoun [Pn]. Several languages distinguish between intimate and casual possession or between kin terms vs. all other possessed nouns. The first type of possession in each case is marked by in several groups by prefixing or preposing a pronoun marker directly to the noun (Pom, Chi, Cho). The second type of possession is often marked by Pn-objective case # N (Pom, Yum). 'Objective case' is variously accusative, genitive, and benefactive in the various languages. When G is an N or NP it is preposed to the possessed N (in Pom, Chi, Sal, Yum, Ser). The possessor N(P) is case-marked objective (in Pom, Yum) and the possessed N is marked to agree for person of possessor (in Chi, Sal, Ser).

- (6) G-obj/ben # N [Pom]
- G # N-his [Chi]
- G #his-N [Sal, Ser]
- Pn-acc # N [Yum]
- Pn-N [Pom (kin terms only; otherwise like G ...N), Sal, Ser]
- Pn:Erg-class-poss-N [Cho]
- Pn:Neu-N [Chi (intimate possession)]
- N-Pn:Act [Chi (casual possession)]

Proto-Hokan probably had postpositions, judging from the evidence of Pomo, Yuman, and Seri. These morphemes are not obviously related to nouns in most present-day languages (although they are in Seri), but their nominal origin is often apparent in etymologies that span the stock.

With the exception of the NA order, all of the word-order traits discussed above, as well as the positioning of the TAM and case markers, are shared by proto-Yuta-Nawan, and – for all I know – proto-Penutian. These facts should not be taken as supporting either diffusion or genetic relationship between the three stocks; although either or both might be the case, they are not necessarily involved; these are morphosyntactic phenomena that are quite typical of languages with OV syntax, and are found as well in Eurasia (e.g., Turkic family) and South America (e.g. Quechwa [Quechua] family). On the other hand, incorporated instrumental pre-pounds and directional postpounds are entrenched in Hokan, but sporadic in Yuta-Nawan (instrumental prefixes in Numic family only) and Penutian (instrumental prefixes and directional suffixes in Maidu, Klamath, and Sahaptian – hardly the 'Penutian kernel').

Viability

In pre-Columbian times most Hokan-speaking populations were nonagricultural, communities were small, and the total population for each language was under 10 000. At present most Hokan languages are dwindling (obsolescent – not being learned by children) or dying (moribund – spoken only by elderly people); several are dead already (marked with % in the classification). Since 1500, Tols and Chontals have become agriculturalists, but Tol is obsolescent, and few children are learning Chontal.

Bibliography

- Dixon R B (1905). 'The Shasta-Achomawi: a new linguistic stock with four new dialects.' *American Anthropologist* 7, 213–217.

- Dixon R B (1906b). 'Linguistic relationships within the Shasta-Achumawi stock.' *Proceedings of the International Congress of Americanists* 15, 255–263.
- Dixon R B (1910). *University of California Publications in American Archeology and Ethnology* 5: *The Chimariko Indians and Language*.
- Dixon R B & Kroeber A L (1913a). 'Relationship of the Indian languages of California.' *Science* 37, 225.
- Dixon R B & Kroeber A L (1913b). 'New linguistic families in California.' *American Anthropologist* 15, 647–655.
- Dixon R B & Kroeber A L (1919). *University of California Publications in American Archeology and Ethnology* 16: *Linguistic families of California*. [on pp. 108–112. Sapir gives evidence for Washo as a Hokan language].
- Greenberg J H & Swadesh M (1953). 'Jicaque as a Hokan language.' *International Journal of American Linguistics* 19, 216–222.
- Grey Crawford J (1973). 'A comparison of Pomoan and Yuman.' M.A. thesis: Idaho State University.
- Grey Crawford J (1976). 'Seri and Yuman.' *Hokan Studies* 305–324.
- Gursky K-H (1964). 'The linguistic position of the Quinigua Indians.' *International Journal of American Linguistics* 30, 325–327.
- Gursky K-H (1965a). 'Ein lexikalischer Vergleich der Algonkin-Golf und Hoka-Subtiaba Sprachen.' *Orbis* 14, 160–215.
- Gursky K-H (1965b). 'Das Proto-Hoka Wort für "Kaninchen".'
- Gursky K-H (1966a). 'On the historical position of Waikuri.' *International Journal of American Linguistics* 32, 41–45.
- Gursky K-H (1966b). 'Ein Vergleich der grammatischen Morpheme der Golf-Sprachen und der Hoka-Subtiaba-Sprachen.' *Orbis* 15, 511–537.
- Gursky K-H (1968). 'Gulf and Hokan-Subtiaban: new lexical parallels.' *International Journal of American Linguistics* 34, 21–41.
- Gursky K-H (1974). 'Der Hoka-Sprachstamm: Eine Bestandsaufnahme des lexikalischen Beweismaterials.' *Orbis* 23, 170–215.
- Gursky K-H (1988). 'Der Hoka-Sprachstamm: Nachtrag I.' *Abhandlungen der Völkerkundlichen Arbeitsgemeinschaft, Heft* 58, 1–37.
- Halpern A (1946). 'Yuma.' In Hoiyer H (ed.) *Linguistic structures of Native America*. 249–288.
- Harrington J P (1932). *BAE-B* 94: *Tobacco among the Karuk Indians of California*.
- Jacobsen W H Jr (1958). 'Washo and Karok: an approach to comparative Hokan.' *International Journal of American Linguistics* 24, 195–212.
- Jacobsen W H Jr (1967). 'Switch-reference in Hokan-Coahuiltecan.' *Studies in Southwestern Ethnolinguistics*. 238–263.
- Kaufman T (1989). 'A research program for reconstructing proto-Hokan: first gropings.' In DeLancey S (ed.) *Proceedings of the 1988 Hokan-Penutian Workshop*.
- Kaufman T (1989). 'Some hypotheses regarding proto-Hokan grammar.' Presented at Hokan-Penutian Workshop, Tucson, Arizona, 4 July 1989.
- Kaufman T (1990). 'Tlapaneko-Sutiaba. OtoMangue, and Hokan: where Greenberg went wrong. Presented at Language and Prehistory in the Americas: a Conference on the Greenberg Classification.' Boulder, Colorado, 22–25 March 1990.
- Kroeber A L (1915). *University of California Publications in American Archeology and Ethnology* 11: 'Serian, Tequistlatecan, and Hokan.
- Langdon M (1974). *Comparative Hokan-Coahuiltecan Studies*. The Hague: Mouton.
- Langdon M (1976). 'The proto-Yuman vowel system.' In Langdon & Silver (eds.). 129–148.
- Langdon M (1979). 'Some thoughts on Hokan with particular reference to Pomoan and Yuman.' In Campbell L & Mithun M (eds.) *The languages of Native America: a historical and comparative assessment*. Austin: University of Texas Press. 692–649.
- Langdon M (1986). 'Hokan-Siouan revisited.' In Cowan W, Foster M K & Koerner K (eds.) *New perspectives in language, culture, and personality*. Amsterdam: Benjamins. 111–146.
- Langdon M & Silver S (eds.) (1976). *Hokan studies*. The Hague: Mouton.
- Langdon M & Jacobsen W (1996). 'Selected bibliography of recent work relevant to Hokan.' *SCOIL* 9, 131–132.
- Leščiner D R (1989). 'Rekonstrukcija nazvanij častej tela v jazykax sem'i xoka.' In *Lingvističeskaja Rekonstrukcija i Drevnejšaja Istorija Vostoka: Materialy k diskussijam na Meždunarodnoj konferencii (Moskva, 29 maja-ijunja 1989g.)*. Čast' 1. Moscow: Izdatel'stvo Nauka. 159–166.
- McLendon S (1973). *University of California Publications in Linguistics* 71: *Proto-Pomo*. Berkeley: University of California Press.
- Mithun M (1999). *The languages of Native North America*. Cambridge: Cambridge University Press.
- Oswalt R L (1977). 'The Hokan word for "tongue": the Pomo evidence.' *University Museum Studies (Southern Illinois University)* 11, 69–78.
- Powell J W (1891). 'Indian linguistic families of America north of Mexico.' *Bureau of American Ethnology Annual Report* 7, 1–142.
- Sapir E (1917). *University of California Publications in American Archeology and Ethnology* 13: The position of Yana in the Hokan stock.
- Sapir E (1920a). 'The Hokan and Coahuiltecan languages.' *IJAL* 1, 280–290.
- Sapir E (1920b). 'A note on the first person plural in Chimariko.' *International Journal of American Linguistics* 1, 291–294.
- Sapir E (1921). 'A supplementary note on Salinan and Washo.' *International Journal of American Linguistics* 2, 68–72.
- Sapir E (1925). 'The Hokan affinity of Subtiaba in Nicaragua.' *American Anthropologist* 27, 402–435, 491–527.
- Swanton J R (1915). 'Linguistic position of the tribes of Southern Texas and Northeastern Mexico.' *American Anthropologist* 17, 17–40.
- Wares A C (1968). *A comparative study of Yuman consonantism*. The Hague: Mouton.

Hopi

K C Hill, University of Arizona, Tucson, AZ, USA

© 2006 Elsevier Ltd. All rights reserved.

Hopi is a Uto-Aztecan language of northeastern Arizona spoken by about 5000 people. Hopi culture focuses a rich ceremonial life on arid land corn (maize) cultivation.

Hopi orthography, which has become semi-official since the publication of the *Hopi dictionary* (Hopi Dictionary Project 1998), uses letter values close to those of the International Phonetic Alphabet but with some exceptions. The palatal glide is orthographic *y*. The high back unrounded vowel is orthographic *u*. Various letter combinations, such as *kw*, *ky*, *ng*, *ngw*, *ngy*, *qw*, *ts*, represent unitary sounds. Long vowels are written double. The apostrophe represents the glottal stop and is not written in word-initial position. Word stress is largely predictable, and only exceptional stress is represented orthographically. Unmarked stress is on the first syllable of disyllabic words and on the syllable containing the second mora of longer words, in other words, it is on the first syllable if it is long, i.e., has a long (double) vowel or ends in a consonant, and on the second syllable if the first syllable is short.

Hopi belongs to the Northern group of Uto-Aztecan languages. A diagnostic sound change for Northern Uto-Aztecan (NUA) is the development of Proto-Uto-Aztecan medial affricate *c to NUA -y-, cf. the words for 'moon', Hopi *muuya*, Nahuatl *mētztli*.

Because of the influence of the writings of B. L. Whorf, Hopi achieved a notoriety as a "timeless language" (Carroll, 1956: 216). As a response to this notion, Malotki devoted a large monograph (1983) to the demonstration that Hopi has an extensive way of talking about time and things temporal. Even Whorf's central claim that Hopi lacks spatial metaphors for time (Whorf, 1941: 83) does not hold up. The very word *qeni* 'space, room (for)' can be used with the sense 'time':

- (1) Ya pumu-y kiihu-t
Q those-ACC house-ACC
amùu-tsa-ve hiisaq qeni?
them-between-at how.much space
How much space is there between those two houses?
- (2) U-ngem qa qeni.
you-for not space
There's no room for you.

- (3) Ya ùu-pe qa qeni, um nuy
Q you-at not space, you me
mihikqw tutuqayna-ni-qö?
at.night be.teaching-FUT-SUBORD(switch-reference)
Is there adequate time for you to teach me tonight?
- (4) Pas qa qeni; pay kya as
very not space; already POTEN perhaps
ayo'wat santi-t aw qeni-ni.
to.another week-ACC to.it space-FUT
There is no time; maybe next week there will be time.

Nor is counting units of time foreign to Hopi, as seen in the following examples:

- (5) Sunat yaasangwu-y sen hóyokpu-t
twenty year-ACC whether increased.amount-ACC
ang pep yes-kyäakyangw
along.it there live-SUBORD
qa hii-ta aw yuku-ya.
not anything-ACC to:it fix-PL
They stayed there twenty years and maybe more, but they never fixed anything.
- (6) Pam siiva-y angsakis koyolaw-qw
he [his.own] each.time be.stashing.
money- away-SUBOR
ACC
oovi naalö-q yaasangwu-y ang
that's.why four-ACC year-ACC along-it
pam aw a'ni àmti.
he to.it very accumulate
Because he kept stashing away his money each time [that he got paid], after four years it really accumulated.

Typologically, Hopi is a head-final, left-branching language; it has a rigid Subject-Object-Verb structure:

- (7) Taaqa taavo-t niina.
man rabbit-ACC kill
The man killed a/the rabbit.

and subordinate clauses precede the main clause:

- (8) Pam peehu-t sami-t
he some-ACC fresh:corn-ACC
a'ki-qe pu' nima.
pick:corn-SUBORD then go:home
He went home when he had picked some fresh corn.

Exceptions are possible, especially for the sake of emphasis. If a subject or object appears after the verb, or if a subordinate clause occurs after a main clause, it is separated by an intonational break (represented by a comma):

- (9) Qa an'ewakw tsovawta, sinom.
in great number were assembled people
A great many people were assembled.
- (10) Pas sòosovik navotiwtá,
very everywhere be.known,
puma so'qö.
they die.SUBORD(switch-reference)
It is known everywhere that they died.

Nouns inflect for number, case, and possessor. Demonstratives and pronouns inflect for number and case. For number, there are four categories to distinguish: singular (one), dual (two), plural (three or more), and distributive (for various types or in various locations). The singular form is unmarked. The dual form, with suffixed *-t(u-)* or *-m(u-)*, is used almost exclusively for animate nouns; inanimates are construed as singular or dual from context. (Things such as clouds, stars, vehicles, and the wind that seem to move of their own accord, as well as sacred things such as places with shrines and developing ears of corn, are treated as animate.) Noun plurals are formed via suffixation, reduplication, suppletion, or combinations of these. Inanimate noun plurals always involve reduplication. Accusative forms are given in parentheses:

| | | | |
|------------------|-------------|----------|---------------|
| (11) | sing. | du. | pl. |
| person | sino(t) | sino- | sino-m(u-y) |
| | | t(u-y) | |
| civilized person | hopi(t) | hopi- | hopiit(u-y) |
| | | t(u-y) | |
| cloud | oomaw | oomaw- | oo'omawt |
| | (u-y) | t(u-y) | (u-y) |
| man | taaqa(t) | taaqa- | tàataqt(u-y) |
| | | t(u-y) | |
| woman | wùuti(t) | wùuti- | momoyam |
| | | t(u-y) | (u-y) |
| boy | tiyo(t) | tiyo- | tootim(u-y) |
| | | t(u-y) | |
| old one | wu-y | wuu-yo-m | wuuwuyom |
| | (wuu- | (wuu-kw- | (wuu-wu- |
| | kw) | mu-y) | kw-mu-y) |
| little spoon | akùwya(t) | (= sg.) | a-'akùwya(t) |
| Hopi village | Hopii-ki(t) | (= sg.) | Hopi-ki-ki(t) |

Demonstratives and pronouns are identical in the dual and plural:

| | | | | |
|------------|-----------|--------------|-----------|--------------|
| (12) | nom.sing. | nom.du./ pl. | acc.sing. | acc.dl./ pl. |
| this | i' | ima | it | imuy |
| that | pam | puma | put | pumuy |
| that | mi' | mima | mit | mimuy |
| over there | | | | |
| I/we | nu' | itam | nuy | itamuy |
| you | um | uma | ung | umuy |

Verbs also show a range of pluralization types, but with verbs it is often the derivational suffix that undergoes the process:

| | | |
|-----------------|-----------|-----------|
| (13) | sing./du. | pl. |
| raise, perfect | aniwna | aniwna-ya |
| think, consider | wuuwa | wuuwa-ya |
| bury | aama | am-ya |

| | | |
|-----------------------|-----------|---------------|
| look for | heeva | hep-ya |
| be called | maatsiwa | maa-matsiw-ya |
| pop out of the husk | tsayo | tsayòmti |
| wear out' | tsàakwa | tsakwàmti |
| descend | haawi | haani |
| die | mooki | so'a |
| sit, dwell | qatu | yeese |
| arrive | pitu | òki |
| be dancing | wunima | tiiva |
| be singing | taw-lawu | taw-lalwa |
| be grinding | hàakokin- | hàakokin- |
| coarsely | ta | tota |
| have a field | paasa'y- | paasa'y- |
| | ta | yungwa |
| go along picking corn | a'kiti-ma | a'kiti-wisa |
| go to pick corn | a'ki-to | a'ki-wisa |

Dual subjects take the singular form of the verb. Compare the following (*tuwa* is the singular/dual subject form of the verb and *tutwa* the plural subject form; *itam(u-)* 'we' is identical in the dual and plural):

- (14) Nu' kawayot tuwa. I saw a horse.
 Itam kawayot tuwa. We (two) saw a horse.
 Itam kawayot tutwa. We (several) saw a horse.

Transitive verbs may also require a different form of the verb, depending on whether the object is plural. Such verbs are also marked for plurality of subject.

- (15) Nu' itàakawayoy pitsina. I brought our horse in.
 Itam itàakawayoy pitsinaya. We (several) brought our horse in.
 Nu' itàakawayotuy pitsina. I brought our (two) horses in.
 Nu' itàakawaymuy ökina. I brought our (several) horses in.
 Itam itàakawaymuy ökinaya. We (several) brought our (several) horses in.

Verbs are inflected for tense (unmarked, future, habitual) as well as subject number. Further, verbs divide into perfective and imperfective stems. In most contexts, unmarked perfectives are construed as past, while unmarked imperfectives can refer to past or present.

Hopi shows an exuberance of derivational suffixing as well as of compounding and noun incorporation into verbs (Hill, 2003). The incorporated noun may be the object of a transitive verb stem:

- (16) tap- kill a cottontail (or two) (tap- < taavo
 nina 'cottontail', niina 'kill singular/
 dual object')
 tap- kill cottontails (qöya 'kill plural
 qöya object')

or the subject of an intransitive:

- (17) kwits- for smoke to rise up (kwits- <
 wunuptu kwiitsingw 'smoke')
 mori- for beans to get planted all
 'üykürümti over (mori 'beans')
 nup-'iwta for there to be snow on the
 ground (nup- < nuva
 'snow')

Verbs may also incorporate non-objects:

- (18) Nu' i-tümkwivi-y öngaspal-mortoyna.
 I my- brine-
 wild:greens- be:dipping:to:moisten
 ACC

I'm dipping my wild greens in salt water.

Hopi is unusual in that incorporated nouns may be specific in reference rather than generic, as common in most languages that show such incorporation:

- (19) Nu' pakiw-maqto-ni; noqw itam
 I fish-go:hunting:for-FUT; so we
 put enang nöönösa-ni.
 that:ACC in:addition eat:PL-FUT
 I'm going fishing; so we can eat it along with
 other food. (cf. *paakiw* 'fish')

Modifiers of the incorporated noun may appear as objects of the verb:

- (20) Pangqaqw pam naat puuhu-t
 from:there he still new-ACC
 kwilatots-ma.
 commercially:bought.shoes-PROG
 He's coming from there wearing brand-new
 shoes. (*kwilatotsi* 'commercially bought
 shoes')
- (21) Kikmongwi pas wuuhagniiqamu-y mong-'oya.
 village.leader very many-ACC leader-
 put(PL.OBJ)
 The village leader sure chose a lot of leaders.
 (cf. *mongwi* 'leader')

Modifiers may also be incorporated:

- (22) puhu-hom-'oyiwta
 new-ceremonial.cornmeal-be.offered
 for fresh sacred cornmeal to be offered
- (23) hihin-hopii-tuqayta
 slight-Hopi-know
 know or speak a little Hopi
- (24) su'aw-wuko-nup-'iwta
 fairly-large-snow-be
 for there to be a fairly large amount of snow on
 the ground

Hopi distinguishes among types of sources of information. Statements based on the direct knowledge of the speaker are unmarked. Statements based on evidence (but not direct experience) are marked by the inferential particle *kur*; statements based on conjecture use the potential modal *kya*; and statements based on secondhand knowledge or hearsay use the quotative particle *yaw*. *Yaw* is especially prevalent in story narration. These evidential particles are relatively free-floating in a sentence; they may occur anywhere before the verb:

- (25) Isikwi hovàati. My meat spoiled.
 Isikwi kur My meat seems to have
 hovàati. spoiled.
 Isikwi kya (I'm afraid) my meat may
 hovàati. have spoiled.
 Isikwi yaw I hear my meat spoiled.
 hovàati.

Dialectally, Hopi divides into three main varieties, differing slightly in pronunciation and vocabulary but with a high level of mutual intelligibility. A prominent phonetic difference resides in the differential development of a feature orthographically represented by the grave accent, as in *wiuti* 'woman'. In Third Mesa speech, such syllables have falling tone. In First Mesa speech and in the speech of the Second Mesa town of Mishongnovi (cf. Whorf, 1946), they end in aspiration. In the speech of the Second Mesa towns of Shipaulovi and Shungopavi, the grave accent feature disappears, such that these syllables are pronounced in the same way as syllables without the grave accent.

Some forms differ between a male speaking and a female speaking:

- | | | |
|-------------|-------------|------------|
| (26) | feminine | masculine |
| | speaker | speaker |
| very | hin'ur | a'ni |
| thank you | askwalí | kwakwhá(y) |
| good/pretty | nukwangw- | loma- |
| big ones | yangsayoqa | hohoskaya |
| be a large | naavinta | kyaasta |
| number | | |
| it's good | is alí | is ali |
| too much | is ehe'tihi | is tathi |

There exists a rich baby talk vocabulary, used speaking with (or as with) small children. Many baby talk words are phonologically quite distinct from normal speech. An example is *uu'na* 'bite' (non-baby *kuuki*), whose first syllable is pronounced with vocal tension and with the teeth together, a pronunciation not easily handled with the orthography. Sometimes there is no normal speech equivalent: *hòona*, 'dance like a kachina', is perfective; the nearest non-baby equivalent is the imperfective *kakatsina*.

A baby talk term may cover a different range of meaning from anything in normal speech: *Tooto* can refer to any insect (as well as baby animals or birds); there is no adult word that covers insects in general.

There is a seemingly archaic register of song and ritual speech. An example is *oo'oomawutu*, a song form of *oo'omawt* 'clouds', in which a number of phonological processes characteristic of normal speech are suspended: vowel shortening, syncope, and final short vowel deletion. This is reminiscent of the archaicaizing song/poetry register of French, in which the colloquial sound change of dropping the mute *e* is suppressed.

Bibliography

Carroll J B (ed.) (1956). *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. New York: The Technology Press of the Massachusetts Institute of Technology and John Wiley & Sons/London: Chapman & Hall.

Hill K C (2003). 'Hopi denominal and noun-incorporating verbs.' In Barragan L M & Haugen J D (eds.) *MIT working papers in endangered and less familiar languages 5: studies in Uto-Aztecan*. 215–244.

Hopi Dictionary, Project. (1998). *Hopi dictionary / Hopiikwa lavàytutuveni: A Hopi-English dictionary of the Third Mesa dialect*. Tucson: University of Arizona Press. [K C Hill, editor-in-chief.]

Malotki E (1979). *Hopi-Raum: Eine sprachwissenschaftliche Analyse der Raumvorstellungen in der Hopi-Sprache*. Tübinger Beiträge zur Linguistik 81. Tübingen: Gunter Narr.

Malotki E (1983). *Hopi time: A linguistic analysis of the temporal concepts in the Hopi language*. Berlin: Mouton.

Whorf B L (1941). 'The relation of habitual thought and behavior to language.' In Spier L, Hallowell A I & Newman S (eds.) *Language, culture, and personality: essays in memory of Edward Sapir*. Menasha, WI: Sapir Memorial Publication Fund. 75–93.

Whorf B L (1946). 'The Hopi language, Toreva dialect.' In Hoijer H *et al.*, *Linguistic structures of native America*. New York: Viking Fund. 158–183.

Hungarian

K É Kiss, Hungarian Academy of Sciences, Budapest, Hungary

© 2006 Elsevier Ltd. All rights reserved.

History

Hungarian is a member of the Finno-Ugric branch of the Uralic language family. The Finno-Ugric peoples represented a kind of linguistic and areal unity, populating the southwestern slopes of the Ural Mountains, until 2000 B.C. Hungarian emerged from among the Ugric dialects around 1000 B.C. The Hungarian tribes left the Finno-Ugric homeland in 500 A.D. and occupied the territory surrounded by the Carpathian Mountains in 895, where they established a Hungarian Kingdom in 1000.

The first written records are Hungarian fragments in a Greek and a Latin text, dating from 950 and 1055, respectively. The first surviving coherent written Hungarian text originated in 1192–1195.

Hungarian has about 13.5 million native speakers, the largest number of speakers in the Uralic language family, 10 million of which live in Hungary. The Versailles Treaties in 1920 annexed one-third of Hungarian native speakers – together with two-thirds of the territory of Hungary – to neighboring countries. As a consequence, Romania now has

1.6 million indigenous Hungarian speakers; Slovakia has 520 000; Serbia-Montenegro, 290 000; Ukraine, 156 000; Croatia, 16 000; Slovenia, 7000; and Austria, 4000. The number of Hungarian speakers in the United States, Australia, and Western Europe is around one million.

Hungarian is fairly homogeneous areally; the only dialect displaying substantial lexical, phonological, and syntactic differences from standard Hungarian is the easternmost, archaic Csángó dialect in Romania.

Names for the Language

Hungarians call themselves and their language *magyar*. Others refer to them by variants of the ancient tribal name *Onogur* (e.g., *Hungarian*, *ungarisch*, *vengerski*).

Language Description

Phonology

Hungarian has the following 14 vowels: a[ɒ], á[a:], e[ɛ], é[e:], i[i], í[i:], u[u], ú[u:], ü[y], ű[y:], o[o], ó[o:], ö[ø], ő[ø:]. (In each pair of symbols, the first one is the letter denoting the phoneme; the second is its phonetic transcription.) The vowels form pairs differing in length; single or double accents mark

long vowels. Hungarian displays vowel harmony; stems (except some recent borrowings) contain either only front vowels (ö, ő, ü, ú) and neutral vowels (e, é, i, í), or back vowels (a, á, o, ó, u, ú) and neutral vowels. Suffixes having both front-vowel and back-vowel allomorphs also participate in vowel harmony – (e.g., *kert-ész-ünk-kel* garden-er-our-with/*fodr-ász-ünk-kal* style-ist-our-with).

The number of consonant phonemes is 24: p, b, t, d, ty[tʲ], gy[dʲ], k, g, f, v, sz[s], z, s[š], zs[ž], h, c[tʰ], cs[č], dzs[j], m, n, ny[nʲ], l, r, j. The letter combination *ly*, occurring in a small set of words, has the same phonetic value as *j*. Consonants also have long versions, indicated by doubling. Adjacent consonants at morpheme boundaries are subject to assimilation processes, among them voicing assimilation:

dob-t-am [doptam]
throw-PAST-1SG
tép-d [tébd]
tear-IMP.2SG

Word stress falls on the first syllable of words, and phrasal stress falls on the first major category of phrases.

Morphology

Hungarian is an agglutinating language. Nouns are inflected for number and case. There are 18 cases, among them a rich system of adverbial cases denoting various location, goal, and source relations:

ház-ban
house-INESS
'in house'
ház-ba
house-ILLAT
'into house'
ház-ból
house-ELAT
'from house'

Possessed nominals bear a possessedness marker and a morpheme agreeing in person and number with the possessor:

lány-a-i-m-at
daughter-POSS-PL-1SG-ACC
'my daughters'

Verbs are marked for tense (present or past), marked for mood (indicative, imperative/subjunctive, or conditional), and bear an agreement suffix indicating the person and number of the subject.

vár-ok
wait-(PRES)-1SG
vár-t-ál
wait-PAST-2SG

vár-j-on
wait-IMP-3SG
vár-ná-nk
wait-COND-1PL
vár-t-ak vol-na
wait-PAST-3PL AUX-COND
'they would have waited'

Definite objects elicit the object-agreement suffix *-ja/le/i* (*lát-já-tok őt* see-OBJ-2PL him 'you see him'), which in many cases fuses with the subject agreement morpheme, yielding an objective conjugation. Verb stems usually denote processes or states; telic accomplishments and achievements are mostly derived by means of verbal particles prefixed to the verb: *eszik* 'eats' – *meg-eszik* up eats; *megy* 'goes' – *be-megy* in goes. Hungarian lacks passive voice. Infinitives selected by an impersonal matrix predicate have their own dative-marked subject, with which they agree in person and number.

Postpositions also agree in number and person with their pronominal complement:

mi mellett-ünk
we near-1PL
'near us'
ti után-atok
you-after-2PL
'after you'

Syntax

The Hungarian sentence displays a 'Topic Focus V XP' order. The topic, or topics, names the referent(s) that the sentence is about. Any argument can serve as topic:

Jánosnak oda-adta Péter a könyvet
John-DAT PRT gave Peter the book

Péter oda-adta Jánosnak a könyvet
Peter PRT gave John-DAT the book

A könyvet oda-adta Péter Jánosnak
the book-ACC PRT gave Peter John-DAT

The immediately preverbal focus is the prosodically and pragmatically most emphatic constituent. If represented by a definite or a specific indefinite noun phrase, it expresses exhaustive identification. Thus *Péter JÁNOSNAK adta oda a könyvet* means 'It was to John that Peter gave the book'; *A könyvet PÉTER adta oda Jánosnak* means 'It was Peter who gave the book to John'. The postverbal order of arguments is free. Universally quantified phrases such as *mindenki* 'everybody', *minden fiúnak* all boy-DAT 'to all the boys' stand between the topic and the focus. The verb, the focus, or both can be preceded by a negative particle: *Nem JÁNOS nem vett autót* not John not bought car 'It wasn't John

who did not buy a car'. The negative particle triggers negative concord: *Senki nem vett semmit* nobody not bought nothing 'Nobody bought anything'. Interrogative phrases appear in the preverbal focus position (e.g., *János kit szeret?* John whom loves 'Who does John love?').

The noun phrase is head-final. Articles and attributive adjectives do not agree with the head noun. Numerals block the plural marking of the noun: *két piros alma* two read apple.

Hungarian is a pro-drop language. When combined with phonetically null morphemes, the copula is also dropped, cf.: *Éva beteg* Eve sick 'Eve is sick' – *Éva beteg vol-t* Eve sick be-PAST 'Eve was sick', *Beteg vagy-ok* sick be-1SG 'I am sick').

Bibliography

- Kiss K É (2002). *The syntax of Hungarian*. Cambridge: Cambridge University Press.
- Kenesei I, Vago R M & Fenyvesi A (1998). *Hungarian*. London & New York: Routledge.
- Kiefer F (ed.) (1992–2000). *Strukturális magyar nyelvtan* 1–3. Budapest: Akadémiai Kiadó. 1–3.
- Kiefer F (ed.) (2003). *A magyar nyelv kézikönyve*. Budapest: Akadémiai Kiadó.
- Kiefer F & Kiss K É (1994). *The syntactic structure of Hungarian*. New York: Academic Press.
- Kiss J & Pusztai F (eds.) *Magyar nyelvtörténet*. Budapest: Osiris Kiadó.
- Siptar P & Törkenczy M (2000). *The phonology of Hungarian*. Oxford/New York: Oxford University Press.

Hurrian

J G Macqueen, University of Bristol, Bristol, UK

© 1994 Elsevier Ltd. All rights reserved.

Hurrian is an ancient near-Eastern language which was in use during the later third and second millennia BC. Proper names of recognizably Hurrian form appear in northern Mesopotamia and the hills to the east from ca. 2300 BC, and by the beginning of the second millennium Hurrian-speaking peoples had established themselves in small kingdoms across much of that area. By ca. 1500 BC Hurrian was in widespread use across an area which reached as far as the Mediterranean coast in northern Syria, and cultic and ritual texts composed in Hurrian are known from as far afield as central Anatolia.

Hurrian is an agglutinative language. The root normally stands in initial position, and is not in itself either nominal or verbal. Noun forms are indicated by the addition of a stem vowel to the root, and this is followed by derivational and then relational (case) suffixes. These include agentive (subject of a transitive verb when the object is expressed) contrasted with zero-suffix (subject of an intransitive verb; direct object of a transitive verb) – i.e., Hurrian is ergative – ; genitive; dative; directive; comitative; locative; stative. Verbal suffixes indicate tense/aspect, negation, mood, etc., and are not preceded by any stem-vowel. As with the noun, derivational suffixes (e.g., iterative, factitive) come first; these are followed by aspectual/temporal forms – perfective (past tense), imperfective (future tense), and neutral/aspectless (present tense). Then come classmarkers indicating either transitive or

intransitive, negative markers, and agentive (person-) and modal markers. Two series of markers exist for the expression of negation and person in the indicative and nonindicative moods.

A brief sentence illustrating some of these points is: *un-a-lla-an žen-iff-ua*. Here *un* has the root idea of 'coming,' the lack of an aspectual/temporal marker indicates a present action, *-a-*, is an intransitive class-marker, *-lla-* is a 3rd PL marker and *-an* is a connective, while *žen* is the 'brother' root, *-iff-* is a 1st SG pronominal, and *-ua* is a dative indicator. The resultant translation is: 'And they [in the context 'gifts'] are coming for my brother.'

Hurrian seems to have died out not long after 1000 BC. A related language, Urartian, was in use in eastern Anatolia from about 850–600 BC. Urartian is not a direct descendent of Hurrian, but rather a parallel development from a common parent language to be dated at least to the earlier third millennium BC. Increasing study of the modern languages of the eastern Caucasus plausibly suggests that Hurrian and Urartian may be members of that linguistic group (see *Caucasian Languages*).

Bibliography

- Bush F W (1964). *A Grammar of the Hurrian Language*. Ann Arbor, MI: University Microfilms.
- Diakonoff I M (1971). *Hurrisch and Urartäisch*. Munich: Kitzinger.
- Diakonoff I M & Starostin S A (1986). *Hurro-Urartian as an Eastern Caucasian Language*. Munich: Kitzinger.
- Speiser E A (1941). *Introduction to Hurrian*. New Haven, CT: American Oriental Society.

Ijò

K Williamson, University of Port Harcourt, Port Harcourt, Nigeria

© 1994 Elsevier Ltd. All rights reserved.

The name Ijò, often anglicized as Ijaw, refers to a language cluster spoken in the Niger Delta area of Nigeria, by people who recognize a common linguistic and ethnic heritage. In older works, Ijò was considered to constitute a single language. But there is neither mutual intelligibility over the whole area, nor a single accepted standard written form, to justify its treatment as a single language. It is of particular interest because both typologically and genetically it is quite distinct from all its neighbors but one.

The Name 'Ijò'

The name 'Ijò' was first recorded by Europeans with a 'j.' In some parts of the area, however, people refer to themselves and their language as Iẓon (Iẓo, Uẓo). The term 'Ijò' is commonly used for the entire linguistic and ethnic group, and 'Iẓon' for the area where people refer to themselves and their language as Iẓon.

Nigerian orthographic conventions used in language names are as follows:

| | | | |
|---|-----|---|-----|
| o | [ɔ] | u | [ʊ] |
| e | [ɛ] | b | [β] |
| i | [i] | | |

The Classification and Nomenclature of Ijò

Ijò was first classified in the Kwa language group (e.g., Greenberg 1963). Bennett and Sterk (1977), however, showed that by lexicostatistic counts it was quite remote from both Kwa and Benue–Congo languages. It is now generally believed to branch off the Niger–Congo family tree at a much higher level (Bendor-Samuel 1989).

The only language closely related to Ijò is the tiny Defaka language, spoken in one section of Nkoro, the easternmost town in which Ijò is spoken (Jenewari 1983, 1989). Together, Ijò and Defaka are referred to as 'Ijoid.'

Using a revised nomenclature for the classification by Jenewari (1989), Ijò can be divided into west and east. West Ijò consists first of the inland Ijò group, comprising three isolated languages: Biseni, Akita (Okordia), and Tugbeni (Oruma); and second of Iẓon, a large language with a complex dialect situation. The late twentieth-century view of the Iẓon dialects distinguishes between the west Iẓon dialects (about seven, Arogbo being typical) and the central Iẓon dialects. The central Iẓon dialects subdivide into north and south; there are some eleven northwest central dialects, Mein being typical, and three northeast central dialects, Kolokuma being typical. There are some four southwest central dialects, east Olo-diana being typical, and some four southeast central dialects, Ẓumò (Boma) being typical.

East Ijò consists of three languages or dialect clusters: first Nembe–Akaha (Akassa), second KAKIBA (Kalabari–Kiriķe (Okrika)–Iḡani), and third Nkoro. There is no complete break in intelligibility; speakers of Nembe–Akaha, the westernmost east Ijò dialects, communicate with speakers of Ẓumò, the easternmost west Ijò dialect.

Geographical Location and Number of Speakers

Ijò is spoken in Nigeria, in the mangrove swamp and fresh-water areas of the Niger Delta and connected waterways bordering on the Atlantic Ocean, from the east of Rivers State to the east of Ondo State. Speakers are estimated at over a million.

Typological Characteristics of the Group

Ijò is a classical S(ubject)–O(bject)–V(erb) language. Adverbials typically occur between Subject and Object, but for emphasis they can be placed sentence-initially; as after-thoughts, or more generally in some east Ijò dialects, they also occur sentence-finally. Tense and aspect markers normally occur after the verb. Serial verb constructions are very common, only the last verb of the series being marked for tense and aspect. There are a few suffixes, 'verbal

extensions' or 'extensional suffixes,' which add meanings such as causative or reciprocal to the meaning of the verb root.

Qualifiers typically precede the noun. Nominalizations and relative clauses are usually formed by adding a general noun such as *person* or *thing* to the end of the part of the sentence which is being nominalized or relativized; alternatively, a relative marker may introduce a relative clause which would be unduly complex.

Ijò has lost the typical Niger–Congo noun class system. It has, however, developed a natural gender system; in Nembe, plural nouns are marked as human or nonhuman, while singular ones are feminine (female beings), masculine (male beings, animals (including females), and a few classes of objects, such as knives or containers), or neuter (humans of undetermined sex and all objects not classified as masculine). This system is marked in demonstratives, definite articles, and pronouns.

Typically, Ijò dialects have vowel harmony with nine oral and nine nasal vowels, though some have reduced the number. Labio velar stops occur in all dialects; in many dialects they are in contrast with voiced implosives. KAKIBA and inland Ijò have two tones plus downstep, some Iẓon dialects have two

tones without downstep, while many other Iẓon dialects and Nembe–Akaha have pitch–accent systems. The typical root structure is CVCV(CV) (where C = Consonant and V = Vowel). Some words begin with vowels and in KAKIBA also with syllabic nasals, as the result of the loss of initial consonants or the retention of the remnants of old noun class prefixes.

Bibliography

- Bendor-Samuel J (ed.) (1989). *The Niger–Congo Languages*. Lanham, MD: University Press of America.
- Bennett P R & Sterk J P (1977). South Central Niger–Congo: A reclassification. *Studies in African Linguistics* 8, 241–273.
- Greenberg J H (1963). *The Languages of Africa*. The Hague: Mouton.
- Jenewari C E W (1983). *Defaka, Ijò's Closest Linguistic Relative*. Port Harcourt: University of Port Harcourt Press.
- Jenewari C E W (1989). Ijoid. In Bender-Samuel J (ed.) *The Niger–Congo Languages*. New York: University Press of America.
- Williamson K (1965). *A Grammar of the Kolokuma Dialect of Ijò*. West African Language Monographs 2. Cambridge: Cambridge University Press.

Ilocano

C Rubino, Rockville, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Ilocano (Iloko, Ilokano, Samtoy) is an Austronesian language with 8 million speakers whose ancestral homeland is northwest Luzon Island, Philippines. It is the third largest language in the Philippines after Tagalog (the basis of the Philippine national language) and Cebuano (Sugbuanon). Ilocano is the largest member of the Cordilleran language family of Northern Philippine languages. Within the family, Ilocano forms its own branch with no close relatives. Other Cordilleran languages include: the Alta branch, the South Cordilleran languages of Kallahan, Ibaloi, Pangasinan, and Ilongot; the Central Cordilleran languages of Isinai, Ifugao, Balangao, Bontok (Bontoc), Kankanay (Kankanaey), Kalinga, and Itneg; Arta; and the Northern Cordilleran languages that can be subdivided into the Cagayan Valley languages of Gaddang, Itawis (Itawit), Agta, Ibanag, Atta, Yogad, and Isneg (Isnag), and the North East Luzon branch that comprises Paranan and the Dumagat (Agta, Casiguran Dumagat) languages.

The original Ilocano provinces include Ilocos Norte, Ilocos Sur, and La Union, but Ilocanos have migrated extensively and even predominate in many localities in the neighboring provinces of Abra, Pangasinan, Tarlac, Benguet, and Cagayan. In the provinces of Abra and Pangasinan, many of the Ilocano speakers are ethnically Tinguian or Pangasinan, respectively, who have traded in their native tongues for the more prestigious *lingua franca*. There are also large communities of Ilocano speakers in the major urban centers of the United States, most notably in California and Hawaii.

Unlike most of the major languages of the Philippines, dialectal variation in Ilocano is minimal. There are two main dialects, Northern and Southern, easily distinguishable by slight lexical differences, intonation patterns, and the pronunciation of the native phoneme /e/, which is pronounced as the *e* in English *let* in the Northern dialects of Ilocos Norte and parts of Ilocos Sur, and as a high, central-back unrounded vowel [u] in Abra, the southern parts of Ilocos Sur, La Union, Tarlac, and Pangasinan.

Ilocano has 15 native consonantal phonemes, and a glottal fricative used in one native word in the

Table 1 Ilocano consonants

| | ±Voice | Labial | Dental | Alveolar | Palatal | Velar | Glottal |
|------------|--------|--------|--------|------------------|---------|--------|----------|
| Stops | – + | p b | t d | | | k g | – (h) |
| Fricative | – | | | s | | | |
| Affricates | – + | | | ts, /ty/ /dy/ | | | |
| Lateral | + | | | l | | | |
| Tap/trill | + | | | r | | | |
| Glide | + | w | | | y | | |
| Nasal | + | m | | n | | ŋ | |

Table 2 Ilocano vowels

| | Front | Central | Centralized back | Back |
|------|-------|---------|------------------|------|
| High | i | | e | u |
| Mid | (ɛ) | | | (o) |
| Low | | a | | |

southern dialect, *haán* ‘no’, the colloquial variant of *saán*, as well foreign loans (see Table 1). Of the consonants, 14 (all but the glottal stop) may appear geminate in roots; the glottal stop only occurs geminate across morpheme boundaries: *agaC-arak* [agaʔʔa:rak] ‘smelling of alcohol’.

Stops are unaspirated and, in final position, unreleased. The voiceless velar stop is pronounced quite far back and fricates before vowels. Unlike in Tagalog, glottal stop does not phonemically appear word-finally. Glottal stop is not represented orthographically word-initially, and word-medially, it is represented with a hyphen. Syllables have mandatory onsets, so the basic syllable structure of the language is CV(C): *ába* ‘taro’ [ʔá:.ba].

The phonemes /t/, /d/, and /s/ palatalize to [tʃ], [dʒ], and [ʃ] before the palatal glide /y/ or its equivalent (i + vowel), e.g., *siák* ‘I’ [ʃak], *tián* ‘belly’ [tʃan], *idiáy* ‘there’ [idʒay]. Because of many borrowings from English, Spanish, and colloquial Tagalog where these palatal sounds are not complex phoneme sequences, the phonemic status of [tʃ], [dʒ], and [ʃ] is open to debate.

Ilocano has four native vowel phonemes /i, e, a, and u/. The new phonemes /o/ and /ɛ/ are post-Hispanic (only in loanwords) (cf. Table 2). In the northern dialects, the phoneme /e/ is pronounced as /ɛ/, not differentiated from its pronunciation in Spanish loanwords.

The high vowel [u] is lowered considerably in word-final syllables, and is thus usually represented in the orthography, e.g., *ások* ‘my dog’ /á:su = k/.

Sequences of two vowels other than the diphthongs /ia/, /io/, and /ua/ are pronounced as two syllables, with an intervening glottal stop in careful speech, *saán* [sa.ʔan] ‘no’ but: *al-aliá* ‘ghost’ [ʔal.ʔal.ya].

Stress is phonemic, e.g., *siká* ‘you, familiar’ vs. *síka* ‘dysentery’. There are, however, certain environments that attract stress. Stress falls on the last syllable if the penultimate syllable is closed: *paltóg* ‘gun’, *takkí* ‘excrement’, *tig-áb* ‘belch’, *pugtó* ‘guess’. Exceptions to this rule include words of foreign origin or words with a velar nasal coda preceding a final syllable: *bibíngka* ‘rice cake’, *karám̐ba* ‘jar’ (Spanish loan). In native words, stress also falls on the last syllable if the last vowel is preceded by a consonant and glide: *sarunuén* ‘follow’, *aniá* ‘what’.

Orthographic double vowels following two consonants usually take stress on the first vowel, with an intervening glottal stop or syllable boundary, e.g., *kanabraang* [ka.nab.rá:.ʔang] ‘gong’, *kulláaw* [kul. lá:.ʔaw] ‘owl’. Words that include two identical CVC sequences separated by a vowel usually will carry the stress on the vowel separating them: *salísál* ‘compete’, *batíbat* ‘nightmare’. There are, however, a few exceptions: *yakayák* ‘sieve’, and *pidipíd* ‘closely set together’.

Vowels before geminate consonants and in stressed open (CV) syllables are automatically lengthened: *sála* ‘dance’ [sá:.la], *babbai* ‘girls’ [bà:b.bá:.ʔi]. Open reduplicated syllables in roots that contain a vowel sequence also bear secondary stress/lengthening: *na.ka-bà:-ba.in* ‘shameful’.

Like its sister Philippine languages, Ilocano is a head-marking, predicate-initial language. When two nominals appear postpredicately, the agent normally precedes the patient: *P{in}artí ti baró ti kaldíng* (slaughter{PERF. TRANS} ART bachelor ART goat) ‘The bachelor slaughtered the goat’. The initial position in Ilocano syntax is reserved for the predicate so constituents in this position are predicative: *Tabbéd ni Bong*, ‘Bong is stupid’, *alutiút* ‘(it is a) house lizard’. When a noun phrase does precede a predicate for

Table 3 Ilocano voice

| Transitivity | Orientation | Affix | Perfective | Example | Gloss |
|--------------|--------------|---------------------------|-----------------------------|------------------------------------|-------------------------------------|
| Intransitive | Actor | <i>ag-</i> <i>-um-</i> | <i>nag-</i> <i>-imm-</i> | <i>agkatáwa</i> <i>dumakkél</i> | 'to laugh' 'to grow, become big' |
| Detransitive | | <i>mang-</i> | <i>nang-</i> | <i>mangán</i> | 'to eat' |
| Transitive | Patient | <i>-en</i> | <i>-in-</i> | <i>suráten</i> | 'to write something' |
| | Directional | <i>-an</i> | <i>-in-an</i> | <i>surátan</i> | 'to write to someone' |
| | Conveyance | <i>i-</i> | <i>in-; iny-</i> | <i>isúrat</i> | 'to write down' |
| | Benefactive | <i>i-an</i> | <i>in(y)-an</i> | <i>idaítan</i> | 'to sew for someone' |
| | Comitative | <i>ka-</i> | <i>kina-</i> | <i>katugáw</i> | 'to sit with someone; seat mate' |
| | Instrumental | <i>pag-</i> | <i>pinag-</i> | <i>pagíwa</i> | 'to slice with; knife' |

pragmatic reasons, it is preceded by a pause or the predicate marking particle *ket*. The phrase *Napintas ni Alessandra* 'Alessandra is pretty' can appear inverted as: *Ni Alessandra ... napintas* or *Ni Alessandra ket napintas*.

Most syntactic structures follow a head + modifier pattern. Genitives follow their nouns > *ti uken-ko* 'my puppy', *ti uken ni Rafael* 'Rafael's puppy'.

Typical of the native languages of the Philippine archipelago, there is a rigid voice distinction in the verbs whereby the semantic relationship between the verb and the pivot (the syntactically most privileged absolutive argument) is signaled by the verb's derivational morphology (Rubino, 2000).

The various voices in Ilocano are shown in Table 3. Each affix has an infinitival/imperative form and a perfective form. Initial CVC reduplication of the verb is employed for progressive (continual) verbs. So the actor voice verb *agdigos* 'take a bath' inflects as *agdigos* 'take a bath', *nagdigos* 'took a bath', *agdigdigos* 'is taking a bath', *nagdigdigos* 'was taking a bath.' The enclitic *=(n)to* which is realized as *=nto* after vowels and *=to* after consonants denotes future time, e.g., *agdigosto* 'will take a bath', *agdigoskanto* 'you will take a bath'. Other derivational possibilities with the root *digos* include *agindidigos* 'pretend to bathe', *pagdigos* 'used for bathing', *kadigos* 'bathing mate', *panagdigos* 'bathing', *idigos* 'to bathe with (+instrument)', *agpadigos* 'have a bath', *padigosen* 'bathe someone else', *pagdigusan* 'bathing place', etc.

Ilocano also has a potentive mode used for actions that are abilitative, coincidental, involuntary, or accidental. Potentive verbs are formed with the prefixes *ma-ka-* or *na-ka-*, e.g., *na-dungparko ti lugan* 'I accidentally hit the car' vs. *D(in)jungparko ti lugan* 'I hit the car (on purpose)'.

Compared to many other Philippine languages, the Ilocano noun marking system is rather simple, with only two case distinctions, a core case (for the two arguments that appear with a transitive verb or the one argument that appears with an intransitive predicate) and an oblique case for other referents.

Table 4 Ilocano articles

| | Non-personal, singular | Non-personal, plural | Personal, singular | Personal, plural |
|---------|---|-------------------------|-----------------------|---------------------|
| Core | <i>ti</i> (neutral), <i>diay</i> (definite) | <i>dagití</i> | <i>ni</i> | <i>da</i> |
| Oblique | <i>ití</i> | <i>kadagití</i> | <i>kenní</i> | <i>kadá</i> |

As is shown in Table 4, plurality in nouns may be expressed by the article. Most countable nouns may also be pluralized by reduplication, e.g., *lalaki* 'boy' > *lallaki* 'boys', *sabong* 'flower' > *sabsabong* 'flowers', *kailian* 'townmate' > *kakailian* 'townmates'.

Ilocano has six sets of pronouns, which encode eight personal distinctions. There are three first person plural distinctions in the language, dual (you and I), exclusive (we but not you), and inclusive (we and you). The second person plural pronouns may be used to a single address to express politeness.

Independent pronouns are used predicatively (see Table 5). Ergative (genitive) and absolutive pronouns are enclitic; they behave like suffixes that do not attract stress shift, e.g., *Napán-ak idiáy* 'I went there', *sá-ak napán* 'then I went'. Monosyllabic enclitics are usually not immediately segmentable by native speakers and some show allomorphic variation dependent upon phonological environment. After the suffixes *-an* (NOMINALIZER; DIRECTIONAL) and *-en* (PAT), the first and second person ergative enclitics fuse with the final *n* to *-k*, and *-m*, respectively, e.g., *basaek/ basa-en = k(o)/* 'I'll read it'. The first and second ergative enclitics also lose their final vowel after vowels, e.g., *adi-m* 'your younger brother', unless they follow the monosyllabic adverbs *sa* 'then' or *di* 'negation' or precede the adverbial enclitic *=(e)n* 'now, already' in which they maintain their full forms, e.g., *kuarta-k* 'my money' vs. *kuarta-ko-n* 'It's my money now'.

When two enclitic pronouns meet in Ilocano, they fuse in such a way that some agentive distinctions are

Table 5 Ilocano pronouns

| Gloss | Indep | Ergative | Absolutive | Oblique | Independent | Reflexive |
|---------------------|----------------|---------------|---------------|------------------|----------------------------|-----------------|
| 1s | <i>siák</i> | = <i>k(o)</i> | = <i>ak</i> | <i>kaniák</i> | <i>kukuák, bágik</i> | <i>bagík</i> |
| 2s familiar | <i>siká</i> | = <i>m(o)</i> | = <i>ka</i> | <i>kenká</i> | <i>kukuám, bágim</i> | <i>bagím</i> |
| 3s | <i>isú(na)</i> | = <i>na</i> | – | <i>kenkuána</i> | <i>kukuána, bágina</i> | <i>bagína</i> |
| 1 dual incl | <i>data</i> | = <i>ta</i> | = <i>ta</i> | <i>kadatá</i> | <i>kukuáta, bágita</i> | <i>bagíta</i> |
| 1 pl excl | <i>dakamí</i> | = <i>mi</i> | = <i>kamí</i> | <i>kadakamí</i> | <i>kukuámi, bágimi</i> | <i>bagími</i> |
| 1 pl incl | <i>datayó</i> | = <i>tayó</i> | = <i>tayó</i> | <i>kadatayó</i> | <i>kukuátayo, bágitayo</i> | <i>bagítayo</i> |
| 2 pl, (2 sg formal) | <i>dakayó</i> | = <i>yo</i> | = <i>kayo</i> | <i>kadakayó</i> | <i>kukuáyo, bágiyo</i> | <i>bagíyo</i> |
| 3 pl | <i>isúda</i> | = <i>da</i> | = <i>da</i> | <i>kadakuáda</i> | <i>kukuáda, bágida</i> | <i>bagída</i> |

Table 6 Ilocano articles and demonstratives

| Visibility | Range | Article | Demonstrative, core, singular | Demonstrative, core, plural | Demonstrative, oblique, singular | Demonstrative, oblique, plural |
|-------------------|----------|-------------|----------------------------------|--------------------------------|-------------------------------------|-----------------------------------|
| Visible ↓ Neutral | Proximal | <i>toy</i> | <i>daytáy</i> | <i>dagitáy</i> | <i>kadaytáy</i> | <i>kadagitáy</i> |
| | Medial | <i>ta</i> | <i>daytá</i> | <i>dagitá</i> | <i>kadaytá</i> | <i>kadagitá</i> |
| | Distal | <i>diay</i> | <i>daydiáy</i> | <i>dagidiáy</i> | <i>kadaydiáy</i> | <i>kadagidiáy</i> |
| Out of sight | Recent | <i>tay</i> | <i>daytáy</i> | <i>dagitáy</i> | <i>kadaytáy</i> | <i>kadagitáy</i> |
| | Remote | <i>di</i> | <i>daydí</i> | <i>dagidí</i> | <i>kadaydí</i> | <i>kadagidí</i> |

neutralized. Thus *gayyem-nak* may mean both ‘I am your friend’ or ‘I am his/her friend’; *Ay-ayaten-da-ka* can mean ‘They love you’ or ‘We love you’.

Ilocano deictics include spatial/temporal demonstratives (which have abbreviated article forms) and temporal adverbs that mark relative time. The temporal adverbs are *itá* ‘now, today’, *itattá* ‘right now’, *itattáy* ‘just a while ago, immediate past’, *itáy* ‘a while ago, recent past’, and *idí* ‘a while ago, remote past’. Temporals can mark both verb phrases and temporal nouns: *N-ag-paráng idí*. (PF-ACT-appear REM.PST) ‘It appeared a while back’, *idí rabií* (REM.PST night) ‘last night’. There is also a future marker (*in*)*ton(o)* that precedes temporal nouns; it cannot be used as a temporal adverb: *intón bigát* (FUTURE morning) ‘tomorrow’.

The nontemporal Ilocano demonstratives mark three degrees of spatial orientation and two degrees of temporality (see Table 6).

The recent and remote articles and demonstratives are used for referents that are not visible in the speech event. They mark referents that may be dead,

non-actual, or somehow distanced from the speech event. Referents that are recently activated into the consciousness of the speaker may also appear with a nonvisible demonstrative. Compare *N-ag-paráng ni Erning*. ‘Erning appeared/showed up’ vs. *Nagparáng daydi Erning*. ‘The late Erning appeared (as a ghost)’; *Ania ti nágan = mo?* ‘What is your name (*nagan*)’ vs. *Ania tay náganmo [manén]?* ‘What was your name [again], (I used to know it)?’

Bibliography

- Foronda M & Juan A (1972). *Samtoy: essays on Ilocano history and culture*. Manila: United Publishing Co., Inc.
- Jocano F L (1982). *The Ilocanos: an ethnography of family and community life in the Ilocos region*. Diliman, Quezon City: Asian Center, University of the Philippines.
- Rubino C (1997). A reference grammar of Ilocano. Ph.D. diss.: University of California, Santa Barbara.
- Rubino C (2000). *Ilocano dictionary and grammar*. Honolulu: University of Hawaii Press.

Indo-Aryan Languages

S Munshi, University of Texas, Austin, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

Indo-Aryan languages (sometimes also called 'Indic' languages) are spoken in India, Pakistan, Bangladesh, Nepal, Sri Lanka, Bhutan and Maldives – the group of seven countries also known as the South Asian Association of Regional Cooperation (SAARC) countries. More than two-thirds of the total population of South Asia speaks Indo-Aryan languages. Owing to 20th-century migrations, considerable numbers of Indo-Aryan speakers have settled in non-SAARC countries, especially in Europe and North America. Indo-Aryan constitutes the largest language group in the Indo-European language family in terms of numerical weight of speakers (said to be approximately one-fifth of the total world population, the total number of speakers exceeding 700 000 000) as well as in terms of the total number of languages in the family (more than 70, of which 24 or more enjoy official status and have literary history). Some of the major languages in the family, both in terms of number of native speakers and literary history, are Hindi, Urdu, Bengali (known to its native speakers as Bangla), Assamese, Punjabi, Gujarati, Marathi, and Oriya. Sanskrit, in terms of historical significance and role in the development of modern Indo-Aryan languages, notwithstanding the very small number of native speakers (only a few hundred), enjoys a very important position within the Indo-Aryan language family. Romani/Romany (popularly known as the 'Gypsy' language), also a language of Indo-Aryan origin, is spoken by several million people in various countries around the world.

Being a region of enormous linguistic diversity, the sociolinguistic situation of South Asia projects a very complex picture. Indo-Aryan languages are in constant and intimate contact with languages belonging to different language families, viz., Tibeto-Burman, Iranian, Dravidian, and Austroasiatic (Munda) languages. In South Asia, from region to region, the accents, dialects, and languages change. The dividing lines between languages are often distorted by large numbers of overlapping and interwoven dialects. It is extremely rare that speakers use only one language; multilingualism is the norm. Speakers often switch between two or more languages. In the context of multilingualism, English, a foreign language, plays a major role. It has a special status in South Asia as a medium of higher education, often associated with social prestige and power. In India, after 54 years of independence, English continues to be the second

official language along with Hindi and other official languages. As a result of multilingualism, code switching, and borrowing, most Indo-Aryan languages, especially those of wider communication, have been considerably influenced. This has resulted in the emergence of various mixed languages, pidgins, and creoles. The contact situation has also led to gradual loss of native languages among many regional groups of people.

Distribution of Major Indo-Aryan Languages

Broadly speaking, Modern/New Indo-Aryan (NIA) languages fall into four geographical groups: northwestern, southwestern, midlands, and eastern. The northwestern group includes Sindhi, Punjabi, Lahnda/Lahndi, Pahari, Dogri, Kashmiri, and other Dardic languages. The southwestern language group comprises Gujarati, Marathi, Konkani, Maldivian, and Sinhala (the official language of Sri Lanka). Major languages in the eastern group are Bengali, Assamese, and Oriya. The midlands Indo-Aryan group consists of Hindi and its different dialects, Urdu (the official language of Pakistan) and its various dialects, a variety of dialects known as Eastern and Western Hindi, and many other languages. At the level of the colloquial language spontaneously spoken/heard on radio, television, or other media, and in the schools/colleges of northern India and Pakistan, Hindi and Urdu are virtually one language. They share a single, identical grammatical system and most of the vocabulary. What makes them two distinct languages are their separate writing systems, different borrowing strategies, and a very robust language ideology. For political purposes, they are essentially two languages.

In terms of total number of speakers, Hindi ranks first, with over 337 million speakers (Cardona and Jain, 2003: 4). It is considered to be the third to fifth most widely spoken language in the world. The term 'Hindi belt' refers to the regions of India with major concentrations of Hindi speakers, viz., the states of Bihar, Uttar Pradesh, Rajasthan, Haryana, Himachal Pradesh, Delhi, and Madhya Pradesh. Hindi is also widely spoken in Mumbai, with distinct regional peculiarities. In addition to Hindi, several other native regional languages, very close to Hindi in mutual intelligibility, are spoken in the Hindi belt. Urdu is spoken by approximately 43.4 million people in India, 6.4 million people in Pakistan, and 275 000 people in Bangladesh. It serves as the lingua franca among the Muslims of South Asia living in and outside the subcontinent. In several studies of Hindi

dialectology, Urdu has been listed as one of the dialects of 'Hindi,' or Hindustani. Historically, Urdu developed from Khari Boli, or 'Hindavi,' which originated in the Delhi area and was considerably influenced by Kauravi, Hariyanavi, Mevati, eastern Punjabi, and Braj Bhaṣa, which, in turn, have developed from the literary language *Śauraseni Apab^hramśa* of Middle Indo-Aryan.

Other major NIA languages with considerably large number of speakers are Bengali (approximately 69.6 million speakers in India, primarily in West Bengal, and 108.6 million speakers in Bangladesh), Marathi (approximately 62.5 million speakers, mainly spoken in Maharashtra), Gujarati (the language of Gujarat, with approximately 40.67 million speakers), Oriya (the language of Orissa, with approximately 28.06 million speakers), Punjabi (approximately 23.37 million speakers in India, mainly in Punjab but also in several other states, and 50.9 million speakers in Pakistan), Assamese (the language of Assam, with approximately 13.08 million speakers), Sindhi (approximately 9.9 million speakers in Pakistan and 2.12 million speakers in India), Nepali (approximately 2.076 million speakers in India and 9.3 million speakers in Nepal), Konkani (approximately 1.76 million speakers in India) (Cardona and Jain, 2003: 4–5), and Kashmiri (over 4 million speakers in the Jammu & Kashmir state (*Ethnologue*, n.d.).

The NIA languages are primarily spoken in South Asia, except for Romani, the only Indo-Aryan language spoken outside the subcontinent. The Romani language evolved as a result of the migration of a population of mixed ethnic and linguistic backgrounds from different parts of India. The migration, which started in the 1100s and continued into the 14th century, took this mixed population to several countries, viz., the Byzantine Empire (Greece), Serbia, Croatia, Bulgaria, Romania, Hungary, Germany, France, Rome, Spain, Catalonia, Cyprus, Switzerland, Russia, and the United States (a total population of approximately 12 million people around the world). Despite obvious Iranian and European influence, Romani is built on a central Indo-Aryan core. It has several regional dialects, Vlax being considered as the 'standard' dialect by various scholars.

Historical Development

The Indo-Aryan language group is a major branch of the Indo-Iranian language subfamily, constituting the easternmost group within the Indo-European language family. Research on Indo-Aryan languages leading to the discovery of their relationship to the rest of the languages in the Indo-European family represents a major breakthrough in historical and

comparative linguistics. The development of Indo-Aryan languages can be traced back to a continuous span of at least 3500 years (Masica, 1991), although many scholars argue for a much earlier date of origin. The period is broadly divided into three stages: Old Indo-Aryan (c. 1500 to 600 B.C.), Middle Indo-Aryan (c. 600 B.C. to A.D. 1000), and Modern/New Indo-Aryan (A.D. 1000 onward) (Masica, 1991).

Based on their geographical distribution, Middle Indo-Aryan (MIA) dialects are classified into four main groups that developed into the present-day NIA languages: northwestern, southwestern, midlands, and eastern. In terms of phonological changes, the most conservative group consists of the northwestern dialects of the MIA, which have retained many of the Old Indo-Aryan (OIA) features and even some Indo-Iranian ones not attested in OIA. To this group belong languages such as Punjabi, Sindhi, Lahnda, Pahari, and many languages of the Dardic group, including Kashmiri. Some of the most characteristic features with respect to the phonological system of the northwestern group include retention of three sibilants (palatal ś, dental s, and retroflex ṣ) of the OIA. In many languages, these have merged into *s* or *ś* and ś; the distinction between the OIA liquids *l* and *r* has also been maintained only in certain northwestern (Shahbazgarhi and Mansehra) and in western (Girnar) groups, whereas elsewhere, they merged with *l*. The most advanced group is the eastern dialect of MIA, which has undergone the greatest number of changes.

The oldest and the best contemporary records of MIA are the Aśokan inscriptions (3rd century B.C.) in various dialects written in Kharoṣṭhī on rock edicts (e.g., Mansehra and Shahbazgarhi versions, which represent the northwestern dialects). Owing to certain changes peculiar to the dialects, the terms *Pāli* and *Apab^hramśa* have been employed to refer to the Aśokan dialects. *Prakrit* is a general term often used in the context of MIA dialects to refer to the vernacular varieties other than *Pāli* and *Apab^hramśa*. *Pāli*, the language of the Hinayana Buddhist canon and based on a midland dialect (Masica, 1991: 52), and a representative of the early MIA, developed in northern India before 200 B.C. but was produced much later in Sri Lanka, Burma, and Thailand (Masica, 1991: 56). Most advanced literary dialects of the MIA are the *Apab^hramśa* dialects, which form a rich source of literary texts dating back to A.D. 600 and earlier. In opposition to *Pāli*, *Prakrits*, and *Apab^hramśa*, is the more prestigious *Sanskrit* (the term *Sanskrita* means 'adorned/polished/cultured (language)'). No single attested Vedic dialect has been proved to be the predecessor of the Classical Sanskrit.

Indo-Aryan languages represent a source of a rich literary tradition over thousands of years. The first recorded texts, the Vedic hymns, date back to 1500 B.C. Study of grammar occupied a very important position in the ancient Indian educational system. Students had to study Vedas as well as ancillaries to Vedas, the *Vedāṅgas*. The latter included *śikṣa* 'phonetics', *vyākraṇam* 'grammar', *c'andas* 'meter', *nirukta* 'etymology' (i.e., explanation of difficult Vedic words), *jyotiṣam* 'astronomy' (the Vedic calendar), and *kalpaḥ* 'ceremonial' (i.e., prescribing the rituals and laying down the procedures for carrying out sacrifices and other ceremonies) (Subrahmaniam, 1999: 1). Extensive information on the details of OIA of different time periods and areas is available in Vedic recitation in *Prātiśākhya* works and in the classic works of ancient Indian grammarians such as Pāṇini (between the 6th and 4th centuries B.C.) and Patanjali (2nd century B.C.) (Cardona and Jain, 2003: 7). Pāṇini's *Aṣṭadhyāyī* is an elaborate treatise on Sanskrit grammar consisting of eight (*aṣṭa*) chapters (*adhyāya*). There is a total of 4000 *sūtras*, which are algebraic-formula-like grammatical statements and rules traditionally assigned to six types: *saṃjña* (definitions of technical terms), *paribhāṣā* (interpretation of grammatical statements), *adhikāra* (defining the scope of a grammatical rule), *vidhi* (signifying operational rules of grammar), *niyama* (restricting the scope of *vidhi* *sūtras*), and *atideśa* (extending the scope of *vidhi* *sūtras*) (Singh, 1991: 1).

Writing Systems

Two very popular ancient Indo-Aryan writing systems are Kharoṣṭhī and Brāhmi. Whereas Kharoṣṭhī is written from right to left and was widely used in northwest India and Central Asia, Brāhmi is written from left to right and was popular elsewhere. Kharoṣṭhī was in use from the Aśokan period until around the 4th century A.D., and is generally agreed to be a derivation of Aramaic. Several theories have been put forth for the origin of Brāhmi. Although Indian scholars have argued for the indigenous origin of Brāhmi, many Western scholars have assumed Brāhmi to be derived from a Semitic prototype (Masica, 1991: 133). Some archaeologists argue for the emergence of Brāhmi in the 5th century B.C., but with the discovery of the Indus Valley (Harappan) civilization of c. 2500–1800 B.C. in the early 20th century, with its still undeciphered script, a new dimension was added to the question of the actual time of origin for the Brāhmi script. Many scholars now believe Brāhmi to be a derivation of the Harappan script, although some disagree, considering this to be a radical viewpoint.

Most modern Indo-Aryan writing systems have emerged from Brāhmi. The best known contemporary and widely used script that originated from Brāhmi is the Devanāgri/Nāgri script. Devanāgri, in various modified forms, is used to write several languages, including Sanskrit, Hindi, Bengali, Marathi, and Nepali. It is an outcome of several developmental stages following Brāhmi. Shāradā, a close relative of Brāhmi, which was used for writing Kashmiri several centuries ago, has been recently argued by some scholars to be the immediate predecessor of Devanāgri, which is built on the same system and corresponds with Shāradā letter for letter, although the letters have considerably changed in form. Shāradā is closely associated with the Takri alphabet used for writing Punjabi.

Modified forms of Perso-Arabic (Nastālīq) are currently used to write Urdu, Sindhi, Punjabi (Gurmukhi and Nastālīq are the officially recognized scripts for writing Punjabi), Kashmiri, Shina, and Khowar. Perso-Arabic writing systems for Urdu and Kashmiri are officially recognized whereas those for Shina and Khowar are still struggling for recognition. Representation of additional sounds not present in Persian/Arabic is achieved by use of additional diacritics to suit the specific phonemic inventories of different languages. Redundant Arabic graphemes are retained only in the spelling of Arabic or Persian borrowings, as are certain redundant letters of neo-Brāhmi (Masica, 1991: 151). Representation of Indo-Aryan vowels in Perso-Arabic script, however, is slightly problematic, especially in languages with large vowel inventories. Short vowels are not represented in this system. Reading, therefore, is often a stressful exercise for nonnative speakers and sometimes even for the native speakers beginning to read a language. It is difficult to distinguish between /e:/, /ai/, and /i:/ in word-medial position because only one symbol is used to represent all three of these vowels. Similarly, there is only one symbol representing the vowels /u:/, /o:/, and /au/ in postconsonantal position. Official script for Kashmiri is more advanced in this respect. Each vowel of the 16-vowel system is represented differently with the help of particular diacritics that are commonly used in literature and are officially recognized.

In the Maldive Islands, a script called Tana/Thaana is used. This script is argued to be phonologically quite efficient, and, written from right to left, is a complete innovation. Tana employs symbols, some of which are based on Arabic diacritics and numerals (Masica, 1991: 152). Konkani is probably the only language that seriously uses Roman script for writing. Representation of peculiar sounds is achieved by use of diacritics and special writing conventions.

Phonological Characteristics

Various phonetic and phonological changes characteristic of the Indo-Aryan group distinguish it from the Iranian group of the Indo-Iranian language family. These include (1) reduction of final consonant clusters; (2) absence of voiced sibilants (except in certain NIA languages/dialects, for which voiced sibilants resulted at a later stage of development); (3) retention of voiced aspirated stop consonants, thus having a fourfold distinction of stop consonants in terms of manner of articulation (except in the northwestern languages, which were influenced by contact with Iranian languages); (4) retention of /l/ vs. /r/ distinction in Classical Sanskrit and many central and (north-) western languages and dialects of MIA and NIA (In Iranian languages, /l/ and /r/ indiscriminately merged into /r/); and (5) development of retroflex consonants, a possible result of Dravidian influence on Indo-Aryan languages and an innovation in the Indo-European languages (Note that Burushaski, a language isolate spoken in the region between South and Central Asia, also possesses retroflex consonants). OIA typically consists of a large number of two-consonant and three-consonant clusters that occur at initial, medial, and final positions. However, a large number of these clusters are drastically limited in MIA by several phonological operations, including epenthesis or assimilation in place and manner of articulation. OIA final consonants were generally lost in MIA dialects, with the exception of *m*, which developed into its vocalic counterpart *ṁ*, as in Pāli *puttā* from Sanskrit *putrāt* 'son' (ablative singular) and *putrās* 'son' (nominal singular), but Pāli *puttām* from Sanskrit *putram* (Cardona, 1987: 441). Dissimilar consonants were assimilated in interior clusters by progressive or regressive assimilation, depending on the nature of consonants involved. Examples are Pāli *puttā*, *sat^hi-*, and *vagga-* from Sanskrit *putrā*-son', *sa^ht^hi-* 'thigh', and *varga-* 'group' (Cardona, 1987: 441). Another MIA development was fricative weakening or lenition (except in some dialects), primarily in the postconsonantal position (i.e., *s* > *h*/C₋; e.g., Pāli *b^hikk^hu* from OIA *b^hikṣu* 'monk'; the process is accompanied by consonantal assimilation), and also in pre-consonantal position, when weakening is accompanied by metathesis and gemination (e.g., Pāli *sukk^ha* from OIA *śuṣka*, with intermediate stages of *sukṣa* and *sukha*, respectively, in which the series of changes involve metathesis followed by fricative weakening and gemination, respectively; and Pāli *pupp^ha* from OIA *puṣpa* via intermediate stages of *puhpa* and *pupha*, in which the series of changes involve fricative weakening followed by metathesis and gemination, respectively) (Bubenik, 1996: 46). In the later stages of MIA, the

OIA geminates that were created as a result of consonantal assimilation (preserved in some dialects, e.g., Pāli) were degeminated. This was accompanied by compensatory lengthening of the preceding short vowel (e.g., OIA *paśyati* 'sees', *śiṣya* 'disciple' > Pāli *passati*, *sisṣa*, and Ardha Magadhi *pāsai*, *sīsa*), (Bubenik, 1996: 28). Compensatory vowel lengthening is also observed in several NIA languages in which vowel lengthening follows the degemination of MIA geminates, as in Hindi *sāt*, Bengali *śāt*, and Marathi *sāt* from MIA *satta*, resulting from OIA *sapta* 'seven'. Exceptions are Punjabi, Kashmiri, and some northwestern Indo-Aryan languages (e.g., Kashmiri *sat^h* vs. Hindi *sāt* 'seven'). OIA allowed light, heavy, and overweight syllables of the type VC, VCC, and V:CC, respectively. In MIA, long vowels were generally shortened in overweight syllables (V:>V/_CC). For instance, OIA *gātra* 'limb' and *grīṣma* 'summer' were replaced by MIA *gatta* and *gimha*, respectively (Bubenik, 1996: 29). Another way of eliminating heavy syllables in MIA was through epenthesis (e.g., MIA *sūriya* from Sanskrit *sūrya*). In many MIA dialects, the three sibilants, *ś* (palatal), *s* (dental), and *ṣ* (retroflex), merged into dental *s*. The distinction of three sibilants (and sometimes two) was, however, retained in the northwest dialects.

Although a remarkable stability is observed in the phonological system of NIA *vis-à-vis* that of MIA, and hence, OIA, different phonological operations affect both consonantal and vocalic sounds, thus differentiating the phonological systems of different NIA languages. The basic Indo-Aryan system of stops (which is also that of Sanskrit and the OIA) theoretically involves five distinctive articulatory positions of the tongue, namely, **labial**, **dental**, **retroflex** (or 'cerebral'), **palatal**, and **velar** (for example, /*p t t c k*/). Typical Indo-Aryan stop consonants are distinguished as voiceless vs. voiced (e.g., *p* vs. *b*) consonants, and unaspirated vs. aspirated (e.g., *p* and *b* vs. *p^h* and *b^h*, respectively) consonants. Some of the phonological changes taking place during the MIA stages continued through NIA, but several new developments also occurred. Voiced aspirates were lost in some languages, especially those of the northwest, being replaced by the corresponding voiced unaspirated obstruents (e.g., Kashmiri and Punjabi). In Punjabi, emergence of a tonal system worked as a compensation, whereby the OIA voiced aspirates were devoiced in addition to loss of aspiration. Retroflex consonants were often replaced by corresponding alveolars in many NIA languages, except in certain phonological environments. Some NIA languages and dialects reveal a tendency to replace the Indo-Aryan palatal stop /*c*/ by a 'dental' or alveolar affricate [ts]. This is observed in Nepali,

some dialects of Bengali, certain Rajasthani dialects, Kumauni, many West Pahari dialects, and several others (Masica, 1991: 94). In some languages and dialects, MIA /c/ surfaces as [c] in certain environments (e.g., before front vowels or palatal glide /y/) and as [ts] elsewhere. A further development in certain languages/dialects is a phonemic contrast between /c/ and /ts/, as in Marathi, Konkani, some West Pahari dialects, and Kashmiri. A step further is realized in the Southern Mewari dialect of Rajasthani, the Chittagong dialect of Bengali, and Assamese, in which the [ts] representation of /c/ is replaced by [s], thus, reducing the phonemic inventory. Palatalized consonants have developed in Kashmiri and some Dardic languages. A complete range of five phonemic nasals (labial /m/, dental /n/, retroflex /ŋ/, palatal /ɲ/, and velar /ŋ/) is found in Dogri, Kalasha, Shina, Sindhi, and some other languages (Masica, 1991: 95). However, in most other languages in this group, many of the different articulatory positions of nasals are often phonologically conditioned. The most frequently occurring nasals are /m/ and /n/. Phonemic palatal nasal /ɲ/ is found in Kashmiri, in which it is a result of a phonologically conditioned diachronic change whereby the palatal vowel inducing palatalization was eventually lost. Although most MIA dialects (except certain north-western and some central dialects) had lost the distinction between the OIA liquids /l/ and /r/ (they merged with either /l/ (Eastern dialect) or with /r/), all NIA languages have both /l/ and /r/ sounds. Certain languages, such as Oriya, Gujarati, many varieties of Rajasthani, Bhili, Punjabi, and some dialects of Lahnda, have developed a retroflex version of the lateral l. Retroflex flap [ɭ] has often been taken as an allophonic variant of /d/ in certain environments in many languages and in certain dialects of some languages. In some rural varieties of Kashmiri, /r/ surfaces as retroflex [ɭ] or even as [d] intervocally, whereas in Standard Kashmiri, the same phoneme surfaces as [r] in all phonological environments (e.g., [ko:ɽi]/[ko:dɽi] vs. [ko:ri] 'girls'). Most Eastern NIA (Magadhan) languages have lost the phonemic contrast between *ś* and *s*, but in some traces these are still seen in free variations of [ʃ] and [s]. A two-way distinction of *ś* vs. *s* is found in certain languages only in specific environments. In certain dialects of Maithili, for example, [s] occurs before back vowels and [ʃ] occurs before front vowels, whereas *ś/s* phonemic contrast is observed before central vowels (e.g., *śālu* 'a variety of grain sorghum' vs. *sālu* 'hedgehog') (Masica, 1991: 98). Typically, there are two semivowels in NIA languages – palatal /y/ and labial /v/. In a number of languages, the occurrence of the semivowels is restricted to semipredictable intervocalic glides. Their

position is weakest in eastern Indo-Aryan and strongest in western Indo-Aryan. There is a phonetic as well as an historical difference between the eastern glides, late in origin and sometimes optional, and western preservations of original OIA semivowels. In many NIA languages, /v/ has a distinctive *w*-like allophone occurring before round vowels (the contact between the upper teeth and the inside of the lower lip in *v* is a loose one).

The basic OIA vowel system is a 10-vowel system constituted of *a*, *a:*, *i*, *i:*, *u*, *u:*, *e*, *o*, *ai:*, and *au:*. As a result of various phonological changes during the history of their development, most NIA languages possess vowel systems considerably different from each other as well as from the OIA vowel system. The minimal NIA vowel system consists of six vowels, which falls into two types: the *Oriya type* (/i e a ɔ o u/) and the *Nepali/Marathi type* (/i e; a ə; o u/). The latter vowels also occur in Lamani and Sadani (Masica, 1991: 109). There are differences in vowel quality, height, etc., of the corresponding vowels in different languages. Bengali has a seven-vowel system, with an additional /æ/ added to the Oriya type vowels. Other vowel systems are the eight-vowel systems of Gujarati (/i e ɛ; a ə; ɔ o u/) and Assamese (/i e ɛ a; ɔ o u/); the nine-vowel systems of Dogri, Rudhari, Sairaki, and several West Pahari dialects; and the 10-vowel systems of Hindi and Punjabi (/i i: e ə; a ə; ɔ o u u:/). The Hindi and Punjabi systems are closest to OIA and are thus considered to be the typical Indo-Aryan vowel system: historically, the OIA diphthongs /ai/ and /au/ were monophthongized into long vowels /æ/ and /ɔ/ at a much later stage of development and are often represented as /ai/ and /au/ (*ai* and *au* are retained in some dialects). Additional vowel systems include the 11-vowel systems of Padari, Bhadrawahi, Kumauni, and Konkani, with varying modifications, and the 12- to 13-vowel systems of Braj, Bundeli, some dialects of West Pahari, and Bashkarik (Masica, 1991: 112). There are languages with even larger vowel systems. For example, Kashmiri has a 16-vowel system consisting of front vowels /i i: e e: ɛ/, central vowels /i i: ə ə: a a:/, and back vowels /u u: o o: ɔ/, with three contrasts in height (high, mid, and low). Up to 20 vowels are found in some Dardic languages. Most NIA languages have nasalized vowels. These vowels are predictable in some languages (in the vicinity of nasal consonants) but in others they are contrastive/phonemic in nature. Certain languages have also developed a tonal system, such as Punjabi (prosodic tone), Lahnda, Dogri, some West Pahari dialects (Khalashi, Kochi, Rudhari), some Dardic languages (e.g., Khawar, Shina, Gawarbat), and Dacca Bengali. Stress in NIA is generally predictable and the position of stress is fixed, although stress patterns

may considerably differ from language to language in terms of complexity of rules/constraints determining stress.

Morphosyntax

Old Indo-Aryan verb morphology exhibits tendencies toward simplification. The aspectual system is much reduced. Within the tense system, the distinction between formal perfect and imperfect is eliminated, leading to a two-way contrast in the preterit from the original three-way contrast. The modal system also went through simplification, so that the subjunctive and indicative gradually were eliminated (Cardona and Jain, 2003: 11). Despite such tendencies, the OIA verb system is very rich as compared to the MIA and NIA systems.

During the development of MIA, there was a general leveling down of the rich OIA morphological system by means of various analogical extensions and operations. Contrast between OIA active and medio-passive was lost, as was the contrast between two kinds of future marking in later stages of MIA. Also, distinction among aorist, perfect, and imperfect was generally eliminated. Productive preterit is provided by sigmatic aorist (Cardona, 1987: 444). Because no final consonants are found in MIA, a reduction in the rich declensional system of OIA consonant stems gives rise to a vowel-ending type only. OIA nominals fall into a number of inflectional types, viz., stems ending in *-a*, *-i*, *-u*, *-an*, *-C(onsonant)*, and *-nt*. Although the OIA vocalic stems remain as such, the consonant ending stems are thematized in the MIA dialects (e.g., Pāli *vijju* < *vidyut* is inflected as an *-u* stem; Pāli *b^haranto* < *b^harant-* is inflected as an *-a* stem) (Bubenik, 1996: 72). In *Ard^ha-Magad^hi* (AMg), *nt*-stems are thematized (Bubenik, 1996: 67). Thematic stems are remodeled in the MIA dialects (e.g., Sanskrit *devaḥ* (nominative singular) > Pāli *devo*, Sanskrit *devam* (accusative singular) > Pāli *devā*) (Bubenik, 1996: 68). In case of nominal paradigms, although OIA maintains the distinction of singular, plural, and dual categories, there is a complete loss of the dual category in MIA. This is also evident in most NIA languages.

The extensive case inventory of OIA was considerably reduced in MIA, especially in later stages. The OIA dative merged with the genitive, and in late MIA, there was merger of the instrumental case with the locative (Bubenik, 1996: 69). Further down the lines of development, nominal paradigms dichotomized into direct (nominative, accusative) and indirect cases, which eventually fused together, providing the groundwork for the NIA oblique case (e.g., in *Apab^h-ramśa*). Toward its development into NIA, the MIA

clitics system also dichotomized into one direct and one oblique form. The oblique form is only used with postpositions (e.g., Hindi/Urdu *-ā* 'NOM.SG' vs. *-e* 'OBL.SG'; Hindi/Urdu *-e* 'NOM.PL' vs. *-ō* 'OBL.PL'). In the system of pronominal clitics, the formal distinction between the direct and oblique forms is found only in OIA and Pāli. Such distinction was lost in some dialects of the MIA (Bubenik, 1996: 90). OIA nominals also had a three-way distinction in gender features, viz., masculine, feminine, and neuter. Although Pāli preserved the three-way distinction of gender, many MIA dialects underwent a gradual simplification in the alignment of three genders. A thematic neuter gender marker was a common target for leveling, which was reassigned masculine ('default') gender in *Ard^ha-Magad^hi*, for example. Leveling was achieved by various series of phonological changes in different gender marking endings (e.g., shortening of final long vowels, deletion of final consonants, convergence of various kinds of stems, and other phonological changes). With a few exceptions, such as Gujarati, which retained the OIA distinction of masculine, feminine, and neuter gender, most NIA languages have only masculine and feminine distinctions. In case of nominal declension, some of the forms were lost in MIA. These were replaced by the corresponding pronominal forms, resulting in identical declension for nominals and pronominals, as in OIA *devaḥ* (nominal declension) vs. Pāli *devo* (nominal declension) and *so* (pronominal declension) (Bubenik, 1996: 92). Another morphological change characteristic of MIA is the resegmentation of inherited causatives (Bubenik, 1996: 120).

Different changes in the syntactic system led to the development of an ergative syntax in NIA. In many NIA languages, the perfective is semiergative. In ergative constructions in most NIA languages, agreement between the subject and the verb is blocked. Although typical Indo-Aryan languages show subject-object-verb word order, an exception is Kashmiri, which is a verb-second language, with the inflected verb occupying the second position of the clause.

Bibliography

- Beames J (1970). *A comparative grammar of the modern Aryan languages of India*. New Delhi: Munshiram Manoharlal.
- Bubenik V (1996). *The structure and development of Middle Indo-Aryan dialects*. Delhi: Motilal Banarsidass Publishers Private Limited.
- Cardona G (1987). 'Indo-Aryan languages.' In Comrie B (ed.) *World's major languages*. London and Sydney: Croom Helm. 440–447.
- Cardona G & Jain D (2003a). 'General introduction.' In Cardona G & Jain D (eds.). 1–45.

- Cardona G & Jain D (eds.) (2003b). *The Indo-Aryan languages*. London and New York: Routledge, Taylor and Francis Group.
- Ethnologue (n.d.). 'Kashmiri.' Available at URL: http://www.ethnologue.com/show_language.asp?code=kas. Ethnologue data from *Ethnologue: Languages of the World*, 15th Edition.
- Grierson G A (1919). *The linguistic survey of India*. vol. VIII, part II. Calcutta: Superintendent Government Printing, India.
- Kachru B B (1969). 'Kashmiri and other Dardic Languages.' In Sebeok T A (ed.) *Current trends in linguistics, vol. 5, linguistics in South Asia*. Paris, The Hague: Mouton.
- Masica C P (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Morgenstierne G (1941). 'The Phonology of Kashmiri.' *Acta Orientalia* 19(1), 79–99.
- Singh J D (1991). *Pāṇini: his description of Sanskrit – an analytical study of the Aṣṭadhyāyī*. New Delhi: Munshiram Manoharlal Publishers Pvt. Ltd.
- Subrahmaniam P S (1999). *Pāṇinian linguistics*. Tokyo: Tokyo University of Foreign Studies; Institute for the Study of Languages and Cultures of Asia and Africa.
- Turner R L (1966). *A comparative dictionary of Indo-Aryan languages* (vol. 1.). London, New York: Oxford University Press.

Indo-European Languages

N E Collinge, Cambridge, UK

© 2006 Elsevier Ltd. All rights reserved.

Cross-linguistic comparison of words and phrases had already become fashionable in the 17th and 18th centuries. In 1786 Sir William Jones famously postulated the kinship of Sanskrit, Greek, and Latin (and possibly the Germanic and Celtic languages), a grouping to which in 1813 Thomas Young gave the title 'Indo-European' (IE) (although the Germans long preferred 'Indogermanisch'). In 1816 specific historical studies appeared, some by Raynouard on the Romance languages (descendants of Latin) and others by Bopp, who added Avestan and Lithuanian to Jones's list. From Grimm's 1822 work on, geneticism flourished and until the 1870s remained Sanskritocentric (but see section on analogy). The tally of IE languages came to include two that apparently stood alone (despite their internal dialectal splits) – Albanian and Armenian – and six subgroups: Balto-Slavic, Celtic, Germanic, Indo-Iranian, Italic, and Greek. To these were later added two Asian members: Tocharian in 1893 and Hittite in 1917, the latter together with Palaic and Luwian (for sure) and Carian and Lycian (almost certainly) forming the 'Anatolian' subgroup. The status of one or two tongues is still debated: Lycian, Phrygian, and 'Illyrian.' With all these and their descendants taken into account, the current total of known IE languages is around 140 (Collinge, 1995a).

Comparative Reconstruction of the Indo-European Language Family Tree

Comparative reconstruction of the family began with linguists using the following techniques.

1. Interlingual comparison of the categories and structures of grammar (done first by Bopp) is usually considered central. This sector is by and large the least subject to outside influence, although some internal alterations have to be discounted (for example, the changing of lexical elements into grammatical forms); and shifts of even the basic type of syntax do occur (see later section on typology). Phonology then long took first place (after Rask, 1818; and Grimm, 1822); this is understandable, as speech sounds are anatomically limited and their shifts easier to plot. Now, however, 'lexical diffusion' is recognized: sound changes occur more quickly in words of certain types and frequency and later spreading.
2. August Schleicher was first to regularize the use of the family tree (borrowed from scientists) to display descent-relations. He also insisted that 'proto' forms (the theoretically established items of a no-longer-extant parent language) be recognizable, and he standardized the marking of them with a prefixed asterisk.
3. It has long been realized that a compared language's evidence must first be cleansed by removing those changed items whose shifts are quite internal to that language and whose earlier shape is diagnosable by 'internal reconstruction.'
4. Possibly the most salutary control is that of 'uniformitarianism,' first proposed by James Hutton in a lecture in 1785. This rules, to cite its negative aspect, that hypotheses as to past entities and behavior must not venture outside the range of historically testified examples.
5. In the 1870s the 'Neogrammarians' (see Collinge, 1995a: 204–205) made their contribution, which still merits a mention. They outlawed those comparative statements of item change that ignored

manifold nonconformities. Shifts must be exceptionless (though they accepted analogy as an interfering force; see later section on analogy). This corrective was necessary at the time; since then, many valid accounts of apparently irregular change have been offered, and causation (which they ignored) has been examined.

Revisions to Indo-European Language Theories

Following these developments in the classification of IE languages were a long succession of revisions and discoveries, some helpful, others disturbing, a few still to be finally assessed.

First among the helpful stands the identification by numerous scholars of ‘analogy’ as a cause or steerer of change. More particular were the increasingly precise analyses of the original phonology of proto-Indo-European (PIE). Its velar consonants were found (Ascoli, 1870) to be of varied sorts. Then several scholars realized that a local palatalizing effect, once recognized outside Sanskrit, established /e, a, o/ as the true set of original lower vowels.

Brugmann (1876) noted how ‘sonant’ root consonants (nasals, to which Osthoff added liquids) might within words become vocalized or accompanied by inset vowels. Next, the high vowels /i, u/ were seen as likely to become consonantal /j, w/ in the right context and were in fact ‘semivowels.’

Perhaps most striking was the discovery – following a hint by Saussure – that PIE had possessed some deep guttural sounds (vaguely ‘laryngeals’), which outside Anatolian were themselves lost but were replaced by lengthening of neighboring vowels or the insertion of a vocalic element into the syllable.

Theories of Historical Language Change

The earliest disturbing theory was that of Johannes Schmidt (1872), following an idea of Schuchardt’s, that historical changes moved like waves across a body of water, being possibly wider, of varying strength, and less adjacent than tree-branching suggests. Sapir (1921) recognized ‘drift’ as shifts that are certain to occur independently sooner or later in any language.

Again, arrays of apparently distinct shifts within a language may be reflexes of one more general feature change (‘conspiracy’ theory).

Yet another mechanism for relating shifts and fixing them in time is known as ‘glottochronology.’ Under this theory, divergence of languages from a

common stock is timed by counting the number of words, from a list of shared items, no longer found to be cognate. It has been applied again to IE recently, in attempts to precisely date the origin of IE languages, but it has also long been denounced as untrustworthy, because lexical borrowing is the readiest kind of interlingual interference.

Lastly, disturbing for its effects on the family tree, comes the ‘glottalic’ hypothesis (see Salmons, 1993). This was promoted most keenly by Gamkrelidze and Ivanov (1984) but forcibly rejected by others, notably Baldi (1999). The traditional, but questionable, set of PIE stop consonants has its simple voiced items /b, d, g/ replaced by the glottalized units /p’, t’, k’/ (aspiration being seen as an unstable additive to the buccal stops). This idea does explain the curious rarity of supposed /b/ in PIE (because /p’/ is phonically awkward), but the apparent prevalence across the extant languages of the realization of these glottals as plain voiced forms is troubling. Moreover, Germanic (and Armenian) would then be isolated as retaining the original voicelessness (in contrast to Grimm).

Other Theories

Some suggestions that may assist the comparative IE historian are still to be finally assessed. Promising was Johanna Nichols’s finding (1986) that marking of a grammatical grouping (such as a noun phrase) – whether on the head member or on the dependent or on both or on neither, or a mixture of these – is a conservative feature and so a clue to genetic status.

Indo-European languages are predominantly dependent-marking. There is certainly resistance to change of marking position, so that at least non-relatedness, or late relatedness by geographical movement, may be provable. Time and further research should clarify this.

More in fashion at present is the mathematical-cladistic approach. ‘Cladistics,’ originally a technical term in biology, is the calculus of family relationships based on shared innovations rather than on simply inherited items. It was brought into linguistics first by Hoenigswald (1987). Now the methodology has been computationalized (see Ringe *et al.*, 2002) as a supplement to traditional family tree study. Applied to the tricky problem of IE subgrouping, cladistics has the interesting result of confirming as unusual the behavior and placing of Germanic (see above, on the glottalic theory). Otherwise, the usual tree is re-established, with Anatolian and Tocharian as early departers and the subgroups Greco-Armenian, Balto-Slavic, and Indo-Iranian forming a joint core. Difficulties are still acknowledged.

Remaining Questions Regarding Classification of Indo-European Languages

First, is an agreed-upon IE tree attainable at all? The various subgroups are themselves soundly based and their membership largely agreed upon; but their relative positioning and chronological fixing (for Hoenigswald, their topology and metrics) are yet unsettled. Revised trees are constantly offered.

Second: Does typology help or hinder? Although it might seem that all IE languages should agree in sound-types and syntactical systems, there are notable discrepancies. For example, the contrast between Hittite verbal categories and those of Greek or Latin, or the present syntax of Hindi or Panjabi versus that of Sanskrit. Genetically linked languages do come to differ in grammar, so Ossetic (Ossetin) is now very unlike other Iranian languages generally. Understandable are attempts to trace in IE grammar a syntactic type-change sequence ‘active’→‘ergative’→‘accusative’ and a reversal in part.

Next is the question of whether it is possible to arrive at a precise date and location for the family’s origin. It is doubtful. Russia, the Black Sea region, and even India have been proposed as birthplaces, and fixing a time of birth is most uncertain (see Clackson, 2000: 451). The current fight is between those who favor a starting point 6000 years ago in Eastern Europe and recent adherents of a beginning (based on biological calculation) 8700 years ago in Anatolia (see Gray and Atkinson, 2003).

Fourth, does archeology help? Pointers such as the ‘agricultural dispersal’ process have been used to co-map the movements of IE people and of their languages (Renfrew, 1987; but note his 1999 rethinking). The two developments have rarely shown a firm consensus as to timing or relationships. Firmer clues may lurk; the field is open.

Next, what of the theory of ‘nostraticism’? Holger Pedersen introduced the term ‘nostratic’ in 1903 to cover the establishment of a linguistic super-group that included IE. Enthusiasts followed. Greenberg actually sought to put together all even remotely interaccessible tongues, including Japanese and Inuktitut. Indo-European was linked with South (and even North) Caucasian, Sino-Tibetan, or even Na-Dene language families. The evidence offered is mostly of supposed lexical equations, although Greenberg did use a selection of morphological likenesses. The notion caused much opposition and seems now to be out of favor.

There are some quite opposed concepts of developing linguistic relationships. ‘Convergence’ has been canvassed – that is, likeness in language caused by

areal contact of people rather than (or as well as) by genetic descent (‘divergence’). Trubetzkoy (1939) pressed it for IE (see also Hock, 1986: 491–498), while others posited it for other families (notably Weinreich, 1968). Later a general periodic analysis was espoused (Dixon, 1997), based on the biological notion of ‘punctuated equilibrium.’ This proposes that languages in areal contact during periods of equilibrium suffer the relatively minor diffusion of features each to each, which misleads observers into supposing a family connection; then come periods of punctuation, in which whole language splits do occur. This in-and-out sequence justifies family trees where the size of the evidence defends them (as for IE); it also allows undeniable likenesses between members of different families (IE included).

Definitions of Terms

‘Indo-European’ is the name of the group of tongues deduced by comparative study to be genetically related. The continued suitability of the name may be questioned. Indian evidence is no longer the major source of relevant information it had been, and since the discovery of Tocharian and the Anatolian subgroup, both in Asia, ‘European’ is less essential. (Possibly the title ‘Eurasian’ might be substituted?)

‘Proto-Indo-European’ indicates the purely theoretical form of the ancestral language, which was formulated by reconciling the reflexes perceived in the group’s tongues (‘comparative reconstruction’). Proto-Indo-European is not itself a totally integrated working language but rather a summary of deductions. Its sources are manifold and involve languages themselves that are somewhat historically tangled. Description of PIE may be ‘systematic,’ as by Brugmann (1897–1916), who gave a static and optimistic account of its categories and forms, or ‘realistic,’ as by Gamkrelidze and Ivanov (1984), for whom PIE is unstable and offers a set of stages of development. (For more information on this, see Lehmann, 2003: 245ff.)

Pre-Proto-Indo-European (PPIE) is a title often used to cover all that can be posited about PIE in its earliest pre-diaspora period (Renfrew, 1999: 271, 284). Sometimes it means ‘pre-morphological PIE’ – that is, a specific stage before the fully organized condition of PIE syntax and morphology (Collinge, 1995b: 4). Some writers have preferred to use the terms ‘PIE I,’ ‘PIE II,’ and so forth.

Conclusion

It is noteworthy that a number (almost a hundred) of ‘laws’ have been formulated to register apparently regular shifts, and their environments, in individual

IE languages or sub-groups, or even across the whole family. The scope of these laws (both as to categories and languages) varies widely, and their validity is always subject to challenge.

Bibliography

- Ascoli G (1870). *Corsi di glottologia*. Turin: Loescher.
- Baldi P (1999). *The foundations of Latin*. Berlin & New York: Mouton de Gruyter.
- Bopp F (1816). *Über das conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache*. Frankfurt: Andreäische Buchhandlung. [Revised 1833–1852 with addition of Avestan and Lithuanian; again 1868–1871.]
- Brugmann K (1876). ‘Nasalis sonans in der indogermanischen Grundsprache.’ *Curtius’ Studien* 9, 285–338.
- Brugmann K (1897–1916). *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen* (2nd edn.). Strasbourg: Trübner.
- Clackson J (2000). ‘Time depth in Indo-European.’ In Renfrew C, McMahon A & Trask L (eds.) *Time depth in historical linguistics*. Cambridge: McDonald Institute. 441–454.
- Collinge N E (1995a). ‘Historical linguistics: History.’ In Koerner E F K & Asher R E (eds.) *Concise history of the language sciences*. Oxford: Pergamon/Elsevier. 203–212.
- Collinge N E (1995b). ‘Before PIE: motives for thematism.’ *Folia Linguistica Historica* 16, 3–27.
- Dixon R M W (1997). *The rise and fall of languages*. Cambridge: University Press.
- Gamkrelidze T V & Ivanov V V (1984/1994). *Indoeuropejskij jasyk i indoeuropejtsy*. Nichols J (trans.). Tbilisi: University Press/Berlin: Mouton de Gruyter.
- Gray R D & Atkinson Q D (2003). ‘Language-tree divergence times support the Anatolian theory of Indo-European origin.’ *Nature* 426, 435–439.
- Grimm J (1822). *Deutsche Grammatik* I (revised version). Göttingen: Dieterich.
- Hock H H (1986). *Principles of historical linguistics*. Berlin, New York & Amsterdam: Mouton de Gruyter.
- Hoenigswald H M (1987). ‘Language family trees: topological and metrical.’ In Hoenigswald H M & Weiner L F (eds.) *Biological metaphor and cladistic classification: an interdisciplinary approach*. London: Frances Pinter.
- Lehmann W P (2003). ‘Realism in Indo-European linguistics.’ In Bauer B L M & Pinault G-J (eds.) *Language in time and space*. Berlin & New York: Mouton de Gruyter. 245–255.
- Morpurgo Davies A (1998). *History of linguistics 4: Nineteenth century linguistics*. London: Longman.
- Nichols J (1986). ‘Head-marking and dependent-marking grammar.’ *Language* 62, 56–119.
- Rask R (1818). *Undersøgelse om det gamle Nordiske eller Islandiske sprogs oprindelse*. Copenhagen: Gyldendal.
- Renfrew C (1987). *Archaeology and language: the puzzle of Indo-European origins*. London: Jonathan Cape.
- Renfrew C (1999). ‘Time-depth, convergence theory, and innovation in Proto-Indo-European: “Old Europe” as a Proto-Indo-European linguistic area.’ *Journal of Indo-European Studies* 27, 257–293.
- Ringe D, Warnow T & Taylor A (2002). ‘Indo-European and computational cladistics.’ *Transactions of the Philological Society* 100, 59–129.
- Salmons J C (1993). *JIES monograph 10: The glottalic theory: survey and synthesis*. McLean, VA: Institute for the Study of Man.
- Sapir E (1921). *Language*. New York: Harcourt, Brace.
- Schleicher A (1861–1862). *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar: Böhlau. [4th edn., 1876.]
- Schmidt J (1872). *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.
- Trubetzkoy N S (1939). ‘Gedenken über das Indogermanenproblem.’ *Acta Linguistica* 1, 81–89.
- Weinreich U (1968). *Languages in contact: findings and problems*. The Hague: Mouton.

Indo-Iranian

E Tucker, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

The Indo-Iranian languages are a major subgroup of the Indo-European (IE) language family, comprising on the one hand the languages of IE descent spoken in the Indian subcontinent (Indo-Aryan), on the other hand those spoken in Iran and adjoining regions such as present day Turkey, Tajikistan, and Afghanistan (Iranian). Indo-Aryan and Iranian languages share so many features that differentiate them

from languages belonging to other branches of the IE family that they must have undergone a period of development in common at a prehistoric date. This ancestor language midway between IE and the earliest attested Indo-Aryan and Iranian is normally called Indo-Iranian or Proto-Indo-Iranian.

Thanks to the survival of nearly continuous documentation, the history of both Indo-Aryan and Iranian can be traced back for around three millennia. Anatolian is attested earlier, but Indo-Iranian ranks together with Greek as the branch of the IE family with the longest recorded history. Indo-Iranian

languages have served as the vehicle for high literary cultures and for the fundamental texts of five major religions (Hinduism, Buddhism, Jainism, Zoroastrianism, Manichaeism).

In the historical period, the development of the Indo-Iranian languages is conventionally divided into three stages (Old, Middle, and New) in both subgroups. Comparison between Old Indo-Aryan (OIA or Old Indic: Vedic and Classical Sanskrit) and Old Iranian (OIr: Avestan and Old Persian) make it possible to reconstruct the principal features of prehistoric Indo-Iranian.

The prehistoric split between the two branches of Indo-Iranian has traditionally been associated with the movement of one group of peoples, the Proto Indo-Aryans into the Indian subcontinent. However, a problem is raised by evidence in 14th century B.C. Hittite texts relating to the kingdom of the Mitanni, whose general population spoke Hurrian but whose rulers bore Indo-Aryan names. Divine names such as Varuṇa and Indra in a Hittite-Mitanni treaty, and numerals such as *aika* ‘one’ (Skt. *éka*-; contrast OP *aiva*-, Av. *aēuua*), *satta* ‘seven’ (Skt. *saptá*) and color adjectives *paritannu* and *pinkarannu* (Skt. *palita*- ‘grey’, *piṅgala*- ‘reddish’) in a treatise on horse-training, point to Indo-Aryan rather than undifferentiated Indo-Iranian in the Near East in the second millennium B.C. Documented only during the most recent stage of Indo-Iranian, the Nuristani languages (formerly often called Kafiri), which are spoken in present-day northeastern Afghanistan, also raise problems of classification. They have been explained variously as Iranian (but distant from Pashto), as Indo-Aryan, or as belonging to a separate branch of the Indo-Iranian family.

Indo-Iranian Phonology

The Indo-Iranian vowel system consisted of three short vowels **a, i, u*, their long counterparts **ā, ī, ū*, and a pair of short and long diphthongs **ai, au*, and **āi, āu*. There were also two vocalic liquids **r, l*.

This simple vowel system resulted from the merger of the IE short vowels **e, o, a* > Ind-Ir. **a* (Skt. *mádhu*-, YAv. *mađu*- ‘intoxicating drink’, cf. Grk. *μέθυ*-; Skt. *dadárśa*, Av. *dādarəsa* ‘I have seen’, cf. Grk. *δέδορκα*; Skt. *yájate*, Av. *yazaita* ‘worships’, cf. Grk. *ἄζειται*). However, at the earliest stage of Ind-Ir. the merger was not complete as palatalization occurred before IE **e*. In open syllables, IE **o* > Ind-Ir. **ā* (Brugmann’s Law: Skt. *dāru*-, Av. *dāuru*- ‘wood’, cf. Grk. *δόρυ*) and merged with **ā* < IE **ē, ō*, *ā* (Skt. *rāj*- ‘king’, cf. Lat. *rēx*; Skt. *āsú*, Av. *āsu*- ‘swift’, cf. Grk. *ὠκύ*-; Skt., Av., OP *mātar*- ‘mother’; cf. Lat. *mater*-). Of the IE resonants, which were allophones

of **y, v, n, m, r, l* in the parent language, **i, u* continued unchanged, the vocalic nasals **ŋ, ɱ* > Ind-Ir. **a* (Skt. *saptá*, Av. *hapta* ‘seven’; cf. Lat. *septem*), and the vocalic liquids **r, l* were preserved as vowels only in OIA (Skt. *mṛtā*- ‘dead’, cf. Lat. *mortuus*).

Cases of vocalic hiatus in the earliest OIA and OIr. poetic traditions and the quantities of some final vowels suggest that the inherited IE laryngeal phonemes survived in some form into Indo-Iranian (Ved. trisyllabic *vaata*- ‘wind’ < **vaHata*-). The development of **h₁, *h₂* is complicated (> *ir/ur* before vowels, > *ir/ūr* before consonants in Skt., but > *ar(ə)* in Av.: Skt. *tirás*, Av. *tarō* ‘across’, ppp. *tīrṇá*- ‘crossed’; Skt. ppp. *pūrṇá*-, Av. *parəna*- ‘filled’). Between consonants, an inherited laryngeal sometimes vocalized > *-i*- (Skt., OP *pitā* nom. ‘father’, cf. Grk. *πατήρ*, Lat. *pater*), but *-i-* from this source is of more restricted occurrence in Iranian than in Indic (OAv. nom. *ptā*, dat. *ḡdrōi* ‘father’; OAv. *dug(ə)dar*-, YAv. *duγdar*- ‘daughter’, cf. Skt. *dūhitār*-, Grk. *θυγάτηρ*).

Ablaut alternations inherited from the parent language are often continued directly: Skt. *ásti* ‘is’, *sánti* ‘are’, Av. *asti*, *hənti*, OP *astiy*, *ha(n)tiy*, cf. Lat. *est*, *sunt* (IE **e/zero*). However, there was a major restructuring of ablaut into a system of quantitative contrasts involving three grades:

| lengthened | full | zero |
|--|-----------------------------------|-----------------------------------|
| Skt. <i>rājānam</i> ‘king’ (acc.) | <i>rājan</i> ‘O, king’ (voc) | <i>rājnas</i> ‘of a king’ |
| Skt. <i>cakāra</i> ‘he has done’ | <i>cakara</i> ‘I have done’ | <i>cakrur</i> ‘they have done’ |
| Av. <i>srauuaiieiti</i> ‘makes hear’ (pres.) | <i>sraotā</i> ‘hear!’ (aorist) | <i>sruta</i> - ‘heard’ (ppp.) |

Indo-Iranian had a large number of obstruents: voiced stops, voiced aspirates, voiceless stops, and perhaps also voiceless aspirates:

| | | | |
|-----------|------------|-----------|--------------|
| <i>*b</i> | <i>*bh</i> | <i>*p</i> | <i>*(ph)</i> |
| <i>*d</i> | <i>*dh</i> | <i>*t</i> | <i>*(th)</i> |
| <i>*j</i> | <i>*jh</i> | <i>*c</i> | |
| <i>*g</i> | <i>*gh</i> | <i>*k</i> | <i>*(kh)</i> |

All four series existed in Indo-Aryan, but the voiceless aspirates are variously considered to be inherited from IE, or to have developed from clusters of voiceless stop + laryngeal within Indo-Iranian (Skt. *pṛthú*- ‘broad’ < IE **pl̥tHu*- ‘broad’; Av. nom. *pañtā* < **ponteH*-s, gen. *pañō* < **pñtH*-os ‘path’), or to represent innovations within Indo-Aryan only. OIA also shows a series of retroflex stops (*t̪, th̪, d̪*, *ḍh* and a retroflex nasal), but these are unlikely to have existed in the Indo-Iranian ancestor, as they are not shared by OIr.

The IE labiovelars underwent a very early conditioned split, producing Indo-Iranian palatals

before original front vowels (Skt., Av., OP *-ca* ‘and’, cf. Lat. *-que*; Skt., OP *jīva-* ‘alive’, cf. Lat. *vīvus*; Skt. *hānti*, Av. *jaiṇti* ‘kills’, cf. Hitt. *kuenzi*) and velars elsewhere (Skt. *yákr̥t*, Av. *yākarə* ‘liver’, cf. Grk. ἥπαρ; Skt. *gnā-*, Av. *g(ə)nā-* ‘goddess’, cf. Grk. γυνή; Skt. *gharmā-*, Av. *garəma-* ‘heat’, cf. Lat. *formus*). On the other hand, the IE palatal stops **k̑, g̑, gh̑* probably developed into affricates in Indo-Iranian, and further developed to OIA *ś, j, h*, but to sibilants in most of Iranian (Skt. *śatām*, Av. *satəm* ‘hundred’, cf. Lat. *centum*; Skt. *ājati* ‘drives’, Av. *azaiti*, cf. Lat. *ago*; Skt. *āṃhas-*, Av. *azah-* ‘distress’, cf. Grk. ἄγχω ‘I throttle’).

There was a partial merger of IE **l* with **r*, but not in all varieties of Old Indic (Vedic *raghú-* ‘quick’, Classical Skt. *laghu-*, cf. Grk. ῥαχύ-, Lat. *levis*), nor completely in Iranian (Av. *rayu-* ‘quick’; but NPers. *lištan* ‘to lick’, cf. Grk. λείχω).

IE **s* developed four allophones in Indo-Iranian; **s, *z*, and **š, *ž* after **r, r, u, ū, au, āu, k, i, ī, ai, āi* (the RUKI rule). These sibilants are all continued in OIr. where **š* is also found after **p, b(h)*, and **s > h* before vowels. In OIA **š > retroflex ṣ* but **z, ž (> ṣ)* were lost, the latter with compensatory lengthening of the preceding vowel and retroflexion of the following stop (contrast Av. *mižda-*, Skt. *mīdhá-* ‘reward’, cf. Grk. μισθό-).

Morphology of Nouns and Pronouns

Vedic, Classical Sanskrit, and Avestan show that Indo-Iranian retained the full IE range of eight cases (nominative, accusative, instrumental, dative, ablative, genitive, locative, vocative) and three numbers (singular, dual, plural). The inflections of thematic nouns and adjectives (stems in *-a* <IE**-e/o*) continued those of the IE parent language (masc. sg. **-as, -am, -ā, -āi, -āt, -asya, -ai, -a*, etc.) apart from the remodeled gen. pl. in *-ānām*, and a second nom. pl. in **-āsas*, which was of limited distribution. On the other hand, the inflection of *ā-* stems was remodeled under the influence of *ī-* stems at an Indo-Iranian date (sg. **-ās, -ām, -ayā, -āyāi, -āyās, -āyā(m), -ai*).

Although inflection depended primarily on the type of stem, for non-thematic stems, it also depended to some extent on inherited patterns of accentuation and ablaut: contrast the rare type of closed genitive sg. OAv. *cašmāṅg* (< **cašmán-s*) from *cašman-* ‘eye’ and the more frequent ‘open’ type Ved. *aryamṇás* from *aryamán-* ‘hospitality’. Even when the Vedic evidence indicates that the accent had become fixed on one syllable, the alternations in ablaut between the suffix and the inflection were continued (from Ved. *tákṣan-*, Av. *tašan-* ‘craftsman’: acc. sg. Ved. *tákṣānam*, Av. *tašānam*; gen., dat. Ved. *tákṣṇas, tákṣṇe* Av. *tašnō, tašne*). A parallel distinction between ‘strong’ and

‘weak’ cases characterized the inflection of many types of Indo-Iranian nominal stems.

Indo-Iranian had a large number of demonstrative pronouns, some of which were also employed in anaphoric function (**sá-ltá-*), and as third person pronouns (**a-*, Skt. *asya, asmāi*, etc.; Av. *ahe, ahmāi*, etc.). There were some characteristic pronominal inflections, which probably began to spread to pronominal adjectives in Indo-Iranian: e.g., nom. pl. masc. **-ai* (Skt. *té*, Av. *tōi* ‘those’; Ved. *anyé*, YAv. *āniie*, OP *aniyai-ciy* ‘others’). Some demonstratives combined several stems (Skt. nom. sg. *ayám, iyám, idám*, acc. *imám, imám, idám*, instr. *enā, ayā*, dat. *asmāi, asyāi*; YAv. nom. sg. *aēm, im, imat* ‘this’, etc.). In the personal pronouns, singular and plural were normally based on completely different stems (‘you’: nom. sg. Skt. *tvám*, OAv. *tuuēm*; nom. pl. Skt. *yūyám*, OAv. *yūš, yūžēm*), and in the oblique cases there were two sets of forms, one accented, the other enclitic (‘me’: gen. Skt. *māma*, OAv., OP *manā*, dat. Skt. *māhyam*, OAv. *maibiiā*; enclitic gen.-dat. Skt. *me*, OAv. *mōi*, OP *-mai*y).

Verb Morphology

The Indo-Iranian finite verb system was of considerable complexity because of its large number of intersecting categories, of voice, tense, aspect, and mood. Tense/aspect distinctions were expressed by different stems (present, aorist, perfect, and a seldom attested future), and three sets of personal inflections, which distinguished active and middle voice. The augment *a-* (<IE**e-*) was also prefixed to forms indicating past time.

Within these tense/aspect categories, the morphology of the stems varied, particularly in the present, which served as the basis for both present and imperfect tenses. Many more formal types than the 10 classes recognized for Sanskrit by the Indian grammatical tradition may be reconstructed: for instance, thematic presents built with the IE suffix **-ské-l-sko-* (Skt. *pr̥cchati*, Av. *pərəsaiti* ‘asks’), reduplicated thematic presents (Skt. *tiṣṭhati*, Av. *hištaiti* ‘stands’), athematic root presents where the accent remained fixed on the root (Skt. *vāste*, OAv. *vastē* ‘wears’). In the aorist category, only s-aorists were morphologically distinctive (Skt. *ābhārṣam* ‘I brought’, *āpaviṣṭa* ‘it purified itself’), while the other types, root aorists, *a*-aorists, and reduplicated aorists resembled imperfects. However, in Indo-Iranian individual verbs developed their own individual systems of contrasts.

The perfect, which originally indicated a state (Skt. *jāgāra*, Av. *jaṃāra* ‘is awake’) or the result of a past action (Skt. *tatākṣa*, Av. *tataša* ‘has fashioned’) was regularly characterized by a special set of inflections

and reduplication (an inherited exception is Skt. *veda*, OAv. *vaēdā* ‘I know’).

The moods were indicative, imperative, subjunctive, optative, and the ‘injunctive’. From a formal point of view, the injunctive resembled an unaugmented imperfect or aorist, and its original function was to mention an action without specifying time or mood. An important use of the injunctive was in prohibitions with the particle **mā* (Skt. *mā*, Av. OP *mā*, cf. Grk. *μη*). The imperative was characterized by distinctive inflectional endings, subjunctives, and optatives by morphemes that could be added to any tense/aspect stem (subjunctive **-a-*, giving **-ā-* in thematic stems; optative **-yā-/ī-*, thematic **-ai-*).

The distinction between active and passive was often realized in finite forms merely by the contrast between active and middle endings. However, present stems in *-ya-* and a particular type of intransitive aorist (RV *ádarsī*, *ádṛśran* ‘it, they became visible’) were also employed with passive value (RV *avāci*, OAv. *auuāci* ‘is/has been said’). The most frequent passive form was the ppp. in *-tā-* (< IE **-tó-*), which was destined to become one of the most important elements in the evolution of the verb system in both the Iranian and Indo-Aryan branches.

Syntax

The word order of Indo-Iranian must have been predominantly SOV, but, as the earliest Vedic and Old Avestan poetry shows, there was a large range of possible variations. In prose, genitives and other qualifiers normally preceded the noun that they determined (Skt. *manor jāyā* ‘wife of Manu’, OP *kurauš puça* ‘son of Cyrus’). Enclitic particles, unaccented pronominal forms, etc., occupied second position (Wackernagel’s Law), and a string of particles followed the first word of a sentence, which was frequently a ‘preverb’ as univertation only took place later. According to Vedic evidence, the finite verb was unaccented in main clauses, but regained its accent in subordinate clauses. Subordination was by means of relative adverbs such as **yád* ‘when, since’ (Skt. *yát*, YAv. *yāt*), **yáthā* ‘so that, as’ (Skt. *yáthā*, YAv. *yaθa*), **yádi* ‘if’ (Skt. *yádi*, OP *yadiy*) and the relative

pronoun **yá-* (< IE **yó-*; Skt., Av. *yá-*), which frequently correlated with a demonstrative pronoun in the main clause.

Nominal compounds were a typical feature of Indo-Iranian sentence structure, but they rarely exceeded two members, whereas in Classical Sanskrit multiple members became regular. Exocentric compounds were particularly frequent (Skt. *ugrābāhu-*, Av. *uγrabāhu-* ‘strong-armed’, Skt. *ṣaḍakṣā-*, Av. *xšuuāš.aši-* ‘six-eyed’).

Lexicon

Although there is a substantial amount of inherited IE material common to both branches of Indo-Iranian, it is with respect to their lexicon that Iranian and Indo-Aryan diverge most clearly (a well-known case is OIr. *ātar-* ‘fire’ versus OIA *ágni-*). However, the speakers designated themselves by the same ethnic (Skt. *āryá-*, OP *ariya-*, YAv. *airiia-*), and a large number of cultural and religious terms appear to date from the Indo-Iranian stage, e.g., Skt. *rátha-*, Av. *raθa-* ‘chariot’, Skt. *yajñá-*, Av. *yasna-* ‘worship’, Skt. *hótar-*, Av. *zao-tar-* ‘priest’, Skt. *sóma-*, Av. *haoma-* ‘sacrificial plant’. The latter all have IE etymologies, but there is also a set of such items that belong to common Indo-Iranian, but are unrelated to anything found elsewhere in IE, and may represent prehistoric loans from an unrelated language and culture: *uśig-* ‘priest’, Av. *usig-*, ‘seer’, Skt. *maghá-*, OAv. *maga-* ‘gift, offering’, Skt. *yātu-*, Av. *yātu-* ‘magic(ian)’.

Bibliography

- Burrow T (1973a). *The Sanskrit Language* (3rd edn.). London: Faber and Faber.
- Burrow T (1973b). ‘The Proto-Indoaryans.’ *Journal of the Royal Asiatic Society*. 123–140.
- Sims-Williams N (ed.) (2002). *Indo-Iranian languages and peoples*. Proceedings of the British Academy 116. Oxford: Oxford University Press.
- Schmitt R (ed.) (1989). *Compendium linguarum iranicarum*. Wiesbaden: Reichert.
- Wackernagel J & Debrunner A (1896–1957). *Altindische Grammatik* (3 vols). Göttingen: Vandenhoeck & Ruprecht.

Inupiaq

L D Kaplan, University of Alaska, Fairbanks, AK, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction and Dialectology

Inupiaq (Inupiatun, North Alaskan and Inupiatun, Northwest Alaska) is an Eskimo-Aleut language spoken in Alaska, part of the Inuit branch or Eastern group of Eskimo languages and distinct from the Yupik languages. Extending across the Arctic from Alaska through Canada to Greenland, Inuit varieties differ from each other in significant ways but nowhere is there found a sharp internal break that would constitute a language border, and so they are considered a dialect continuum (**Figure 1**). The Inuit groups of the Eastern Arctic are a diaspora that spread from Alaska, and the Bering Sea area is recognized as the homeland of the Eskimo-Aleut language family and people (*see Eskimo-Aleut*).

Inupiaq comprises two dialect groups, North Alaskan Inupiaq (NAI) and Seward Peninsula Inupiaq (SPI), each with two dialects. NAI includes the North Slope and the Malimiut dialects, with the

former spoken along the Arctic coast and the latter found around Kotzebue Sound, along the Kobuk River and south at the head of Norton Sound. SPI includes the Qawiaraq dialect, found along the northern shore of Norton Sound, on the southeastern Seward Peninsula and at Teller, and the Bering Strait dialect is found along the shores of Bering Strait and on the offshore islands, King Island (now uninhabited), and Little Diomed. Dialects are distinguished primarily in terms of phonology, lexicon, and morphology and include a number of subdialects.

Phonology and Writing

The Alaskan Inupiaq writing system was developed by Roy Ahmaogak, a Barrow Inupiaq minister, and linguist Eugene Nida in 1946 and has undergone revisions since. Current orthographic symbols are given below with equivalent phonetic symbols in parentheses where the two differ. The entire palatal series is absent in SPI, *z* is present only in the Bering Strait dialect, and *e* has limited occurrence, found in SPI only, particularly on Little Diomed Island.

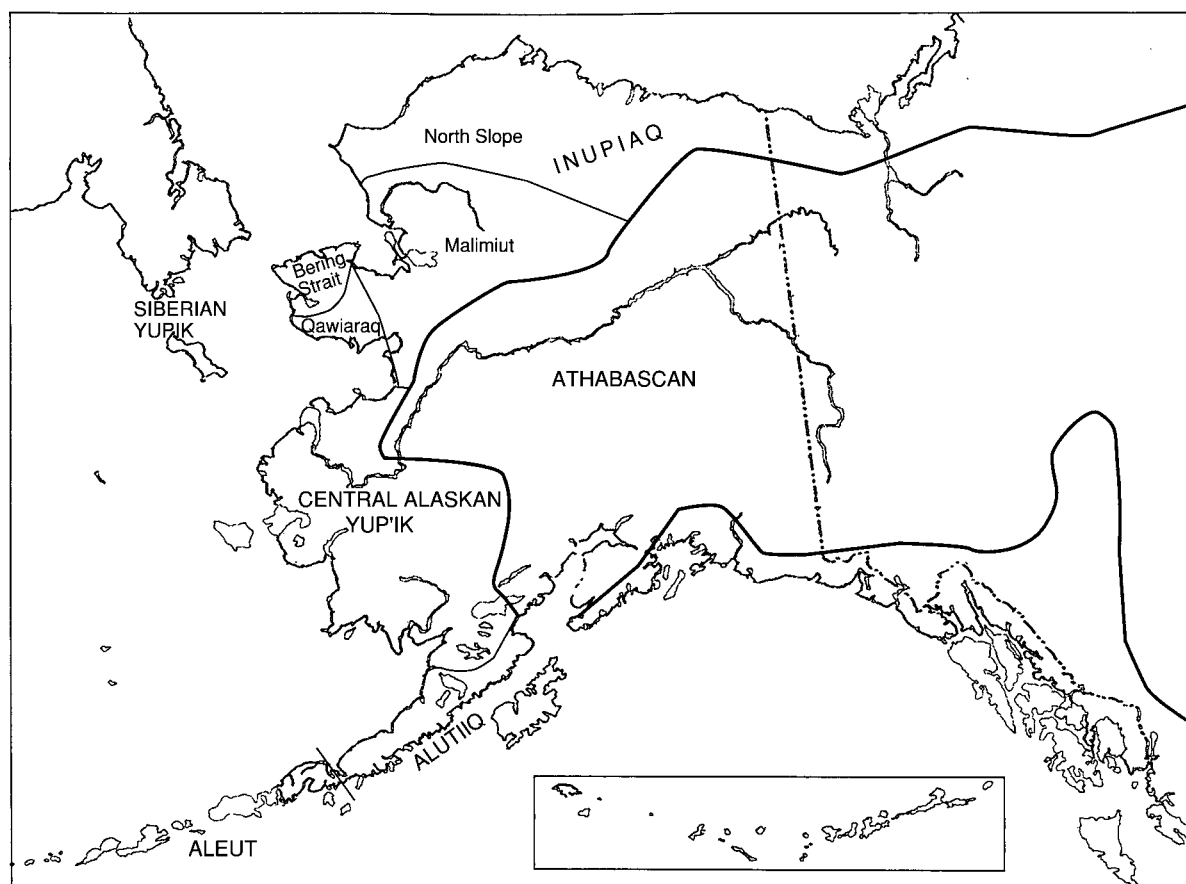


Figure 1 Eskimo-Aleut languages of Alaska.

Consonants:

p t ch(č) k q '(ʔ)
 ɬ(ɬ) ɬ(ɬ) ɬ(l'), y g (ɣ) ǵ(R)
 m n ñ ŋ

Vowels and diphthongs:

i e(ɛ) u
 a

Consonants and vowels can be long and are then written double. Any two vowels with the exception of *e* may be paired: *ai ia ua au in ui*. In most of Malimiut, certain diphthong pairs have coalesced and are pronounced identically: *ia* and *ai* are pronounced [e:], *au* and *ua* are pronounced [o:], and *iu* is pronounced like *ii* as [i:]. A major phonological phenomenon in Inuit languages is consonant assimilation, which increases in magnitude as one travels east (*see West Greenlandic*). Alaskan Inupiaq has differentiated clusters, although the North Slope dialect has assimilation by manner of articulation. Palatalization of alveolar consonants is notable in NAI, with palatalization of some velars found in much of Malimiut. A major distinguishing feature of the Inuit branch is the simplification of the Proto-Eskimo four-vowel system by the merger of the central vowel *ɪ*, principally with *i*. Reflexes of historical *i* trigger progressive palatalization of alveolars and assibilation of prevocalic *t* in NAI: NAI *ĩnuk* 'person' but SPI *inuk* and NAI *isiq-* 'enter' but SPI *itiq-*. Reflexes of the fourth vowel *ɪ* are typically [i] but may undergo deletion or alternation with other vowels.

Morphology and Syntax

Inupiaq is a highly polysynthetic language, and suffixation creates very long words, so that a word is often equivalent to an entire English sentence. Words are typically constructed of a noun or verb stem, one or more derivational suffixes, and an inflectional ending, and may be singular, dual, or plural in number. Examples are from the North Slope dialect:

- (1) atuḡ-nia-ŋit-chuk
 sing-FUT-NEG-3du/INTRANS
 'they (two) will not sing'

Verbs may be intransitive or transitive, and transitive verb endings mark number and person of both subject and object.

- (2) tautuk-kaat
 see-3plSUBJ/3singOBJ
 'they see him/her/it'

The verb ending here expresses the subject and direct object, and pronouns are unnecessary, although they can be added in, as can nouns. Pronouns express

person and number but not gender, which is expressed only by nouns. There are no articles.

- (3) aṇutit aḡnaq tautuk-kaat
 men.pl woman see-3plSUBJ/3singOBJ
 'the men see the woman'

Word order is Subject-Object-Verb, although the complex inflectional system makes free word order possible. Through the influence of English, many speakers now prefer SVO order.

Lexicon and Ethnonyms

Lexical borrowing is primarily from Russian and English. Although the Russians never had a permanent presence in Inupiaq Alaska, the dozen or so Russian borrowings found in Inupiaq were probably introduced through trade with mainland Yupiks to the south and are found principally in SPI and Malimiut. Early English borrowings were introduced beginning in the late 19th century through contact with traders and whalers and were well integrated phonologically (e.g., *palauvak* 'flour'). Modern-day bilingualism often gives rise to use of English words in Inupiaq speech, but these occurrences cannot be considered true borrowings.

Besides indicating the name of the language, 'Inupiaq' can also be used in English either as an adjective (e.g., the Inupiaq language) or a singular noun for a person (e.g., an Inupiaq from Barrow). The plural 'Inupiat' also is used for people (the Inupiat people, the North Slope Inupiat). With consonant palatalization, NAI uses the spellings 'Iñupiaq' and 'Iñupiat,' whereas SPI lacks palatalization and uses 'Inupiaq' and 'Inupiat.' Alaskans also use 'Eskimo,' although this term is disfavored in Canada.

Population and Viability

There are some 13 500 Inupiat in Alaska, about 3000 of whom speak the Native language. Inupiat are now bilingual or speak only English. Most speakers are in their late forties or older, and in some areas the youngest Inupiaq speakers are in their sixties or even seventies. The language shift to English is brought about by a number of factors: government, education, and media are largely in English (although Inupiaq is often heard on the radio); monolingual English speakers have lived among the Inupiat for decades; and airplanes have made travel outside the area easy. In addition, past Inupiaq language use was discouraged and often punished by teachers and school officials. In 1998, a majority voted to make English Alaska's official language, sending a negative message in the view of Native language supporters. The continued use of Inupiaq as a spoken language is

threatened, and efforts at revitalization consist largely of school language classes. Language immersion programs exist in elementary schools in Kotzebue and Barrow, from which it is hoped that new generations of Inupiaq speakers will emerge.

Bibliography

- Berge A & Kaplan L (2005). 'Contact-induced lexical development in Yupik and Inuit languages.' *Etudes Inuit Studies* 29(1–2), 285–305.
- Fortescue M, Jacobson S & Kaplan L (1994). *Comparative Eskimo dictionary with Aleut cognates*. Fairbanks: Alaska Native Language Center.

- Kaplan L D (1981). *Phonological issues in North Alaskan Inupiaq*. Research paper no. 7. Fairbanks: Alaska Native Language Center.
- MacLean E (1980). *Abridged Inupiaq and English dictionary*. Fairbanks: Alaska Native Language Center and Barrow: Inupiat Language Commission.
- MacLean E (1986a). *North Slope Inupiaq Grammar: first year*. Fairbanks: Alaska Native Language Center.
- MacLean E (1986b). *North Slope Inupiaq Grammar second year*. Fairbanks: Alaska Native Language Center.
- Webster D H & Zibell W (1970). *Inupiat Eskimo Dictionary*. Fairbanks: Summer Institute of Linguistics.

Iranian Languages

P O Skjærø, Harvard University, Cambridge, MA

© 2006 Elsevier Ltd. All rights reserved.

Iranian languages have been spoken for 3000 to 4000 years in various parts of southern Russia and the Caucasus, Central Asian republics, Xinjiang, Afghanistan, Pakistan, Iran, Iraq, and Turkey, as well as in the diaspora. The language spoken by the largest number of people today is (New) Persian (*see Persian, Modern*), or Farsi. Iranian languages are closely related to the Indo-Aryan (*see Indo-Aryan Languages*) languages, with which they constitute the Indo-Iranian (*see Indo-Iranian*) subgroup of the Indo-European language family, to which most European languages also belong.

The Iranian languages are known from three main periods, commonly classified as Old, Middle, and New (modern) Iranian. Historically, this division corresponds roughly to (1) the pre-Achaemenid and Achaemenid period (down to ca. 300 B.C.E.), (2) the period down to the Arabic invasion of Iran and the spread of Islam (7th century C.E. in the west, up to 11th and 12th centuries C.E. in the east), and (3) the Islamic period.

Iranian languages are written in a large variety of ancient and modern scripts.

Old Iranian Languages

Avestan (*see Avestan*), the language of the Avesta, was spoken in areas of Central Asia, Afghanistan, and eastern Iran from the 2nd millennium till about 500 B.C.E. Two stages of the language are known, the older of which is comparable to the oldest Indic seen in the

Rigveda and was therefore probably spoken about the mid 2nd millennium B.C.E., while the younger is comparable to Old Persian and was probably spoken ca. 1000–500 B.C.E. The Avestan texts were transmitted orally for 1000 to 2000 years and written down only in about the 6th century C.E.; the earliest manuscripts date from the 13th century. Its phonetic form is therefore that of the latest oral transmitters. The present Avestan corpus is relatively small.

Old Persian (*see Persian, Old*), ancestor of Middle and New Persian, was spoken in southwestern Iran in the first half of the 1st millennium, till about 400 B.C.E. It was the official language of the Achaemenid dynasty and was written in a cuneiform script. The Old Persian corpus is quite small.

Numerous words in the Old Persian inscriptions have a phonetic form different from what is considered to be genuinely Old Persian. It is assumed these are from Median, a language spoken to the north of Old Persian, of which no texts survive.

Middle Iranian Languages

Khotanese (*see Khotanese*), spoken in the kingdom of Khotan in southwestern Xinjiang, is known from a variety of texts dating from about the 6th century to the end of the 10th century C.E. Tumshuqese, spoken in Kucha in northwestern Xinjiang, is known from the same type of sources, but much less well. These two languages were written in the southern and northern variants of Brahmi, respectively.

Sogdian (*see Sogdian*), spoken in Sogdiana, modern Central Asia, is known from texts dating from the 4th to the 10th century. It was written in Sogdian,

Manichean, Syriac, and, occasionally, northern Brahmi scripts.

Chorasmian (*see* **Chorasmian**), spoken in the Chorasmian state along the upper Oxus/Syr Darja as late as the 14th century, is known from inscriptions, coins, and interlinear glosses in two Arabic works, a legal work and a dictionary. It was originally written in Chorasmian, later in Arabic script.

Bactrian (*see* **Bactrian**), spoken in the Greco-Bactrian kingdom founded by soldiers of Alexander the Great in northern Afghanistan, is best known from royal inscriptions and private documents dating from the 2nd to the 8th century. It was written in lapidary and cursive Greek and in Manichean scripts.

Parthian, spoken in the Parthian kingdom to the northeast of the Caspian Sea, is known from inscriptions, letters, economic and legal documents, and Manichean texts dating from about the 1st century B.C.E. to the 10th century. It was written in Parthian and Manichean scripts.

Middle Persian (also Pahlavi; *see* **Pahlavi**), was spoken in southwestern Iran and became the official language of the Sasanian dynasty. It is known from a large variety of texts, notably Zoroastrian texts, dating from about the 1st century B.C.E. to the 13th century, although it had been replaced by modern Persian as a spoken language by about the 8th century. It was written in Middle Persian and Manichean scripts.

New Iranian Languages

This section addresses the main literary languages spoken today, many of them now also in the diaspora, especially in Europe and America.

Various forms of Persian (*see* **Persian, Modern**) are spoken throughout Iran, Afghanistan, Tajikistan, and in adjacent areas. It is written in Arabo-Persian, Cyrillic (Tajik), and Hebrew scripts (Judeo-Persian). Ossetic (ossete) is spoken in Ossetia, in the southern Caucasus, in two main variants, Digoron (the more archaic) and Iron, with subvariants. It is written in the Cyrillic script, except in the south, where Georgian is used.

Kurdish (northern, central, and southern) is spoken in three principal variants in eastern Turkey and Syria, northern Iraq, and western Iran, as well as in surrounding areas. It is written in a modified Persian or standard Turkish Latin script.

Balochi (*see* **Balochi**; or Baluchi, in several dialects) is spoken in eastern Iran and western Pakistan, but also in southern Afghanistan and Central Asia. It is written in the Arabo-Persian script.

Pashto (*see* **Pashto**; several dialect groups) is spoken mainly in Afghanistan and Pakistan. It is written in the Arabo-Persian script.

All the modern languages contain a large number of Arabic words and a smaller number of Turkish words. The easternmost dialects have also borrowed extensively from Indic languages. Differently, Ossetic has borrowed extensively from the neighboring Caucasian languages.

Nonliterary languages and dialects comprise the following:

Northwestern and central Iran: Taleshi on the western shore of the Caspian Sea and Tati dialects from Iranian Azerbaijan through the Central Province and into Gilan; the Caspian dialects (Gilaki in Gilan, Mazanderani, and several dialects on the northern edge of the great salt desert, the Dasht-e Kavir); Gurani (Bajelani) and (including Awromani) in eastern Iraq and western Iran; and Zaza (Dimli) in eastern Turkey; and the Central dialects, comprising a number of more or less interrelated dialects spoken north and south of the Dasht-e Kavir.

Southwestern and southern Iran: Lori (Luri) (in several varieties) and Bakhtiari; Fars dialects; Larestani (in several dialects) in Larestan; dialects in the area from Bandar-e ‘Abbas (Bandari) and Hormoz to Minab and Bashkardi in Bashkerd; and Kumzari on the Musandam peninsula across the Strait of Hormoz (it is unknown whether this is spoken today).

Afghanistan and Central Asia: Parachi and Ormuli (Ormuri) in central Afghanistan and across the border in Pakistan; Ishkashmi and Sanglechi to the west of the Wakhan corridor; Yidgha and Munji/Munjani in eastern Afghanistan and western Pakistan respectively; the Yazghulami-Shughni (Yazgulami-Shugni) group in northern Afghanistan and Central Asia, including Sarikoli in western Xinjiang; Yagnobi (Yagnobi) in the Yagnob valley in Tajikistan; and Wakhi in the Wakhan corridor in northeastern Afghanistan.

Genetic Relationships among the Iranian Languages

Only the Persian languages are descended in a more or less direct line. Of the other dialects, the following are more closely related than others: Yidgha–Munji and Bactrian; Yagnobi and Sogdian; and Wakhi and Khotanese.

Characteristic Phonological Features of Iranian Languages

Phonetic developments differentiating Iranian from its Indo-Iranian ancestor and so also distinguishing

it from Old Indic include the following (see also Indo-Iranian):

1. Indo-Iranian *s* after vowels became Iran. *h* (e.g., OInd. *asura*-, Av. *ahura*- 'lord').
2. Voiced aspirated stops and affricates lost the aspiration (e.g., OInd. *bhara*-, OIran. *bara*- 'carry'; OInd. *dhā*- 'to place' and *dā*- 'to give,' OIran. both *dā*-; Indo-Iran. **j^h an*- 'strike' > OInd. *han*-, Av., OPers. *jan*-).
3. Indo-European palatal velars (*k̂*, *ĝ*, *ǵ^h*) and palatalized velar stops (*k^y*, *g^y*, *ǵ^{yh}*) developed differently in Indic and Iranian (approximate phonetic values: *ś* [ʃ], *ć* [tʃ], *ž* [ʒ], *ǰ* [dʒ], *š* [ʂ], *č* [tʂ], *ž* [ʒ], *ǰ* [dʒ>]):

| IEur. | proto-Ilr. | OInd. | proto-Ir. |
|------------------------------|--|----------------|-------------------|
| <i>*k̂, ĝ, ǵ^h</i> | <i>*ć, f, f^h</i> | <i>ś, ṣ, h</i> | <i>*ć, *f, *f</i> |
| <i>*k, g, g^h</i> | <i>*k^y, g^y, ǵ^{yh}</i> > <i>*č, ž, ǰ^h</i> | <i>č, ṣ, h</i> | <i>*č, *f, *f</i> |

4. The Indo-Iranian unvoiced stops *p*, *t*, *k* became spirants *f*, *θ*, *x* before other consonants, including original laryngeals (*H*) (OInd. *cakra*-, Av. *caxra*- 'wheel'; OInd. *trita*-, Av. *θrita*- 'third'; OInd. *priya*-, Av. *friia*- 'dear'; Indo-Iran. **ratHa*-, OInd. *ratha*-, Av. *raθa*- 'chariot'; but **pt* → **pt* > Av. *pt*: Av. *hapta* 'seven,' NPers. *haft*).
5. The geminated dentals *-tt*-, *-dd^(h)*- (*-t̥t*-, *-d̥d^(h)*-) became *-st*-, *-zd*- (OInd. *vitta*-, Av. *vista*- 'found' < *vid*-; OInd. *addhā* 'truly,' OPers. *azdā* 'well-known').
6. Indo-Iranian laryngeals remained between vowels but were lost between consonants (Iran. **daHah* 'gift,' OAv. *da'o*, spelled *dā*; Indo-Iran. **pHtar*- 'father' > OAv. *ptar*-, OInd. *pitar*-), although a vowel was inserted (or the vocalized laryngeal kept) in initial groups (YAv., OPers. *pitar*-).

Proto-Iranian in turn split into several distinct dialect groups characterized, among other things, by the developments of the palatal affricates *ć*, *ǰ* and the groups *ćw* and *ǰw*.

| Proto-Ir. | SW-Iran. | Central-Iran. | NE-Ir. |
|----------------|-----------------|---------------|-------------|
| <i>*ć, f</i> | <i>θ, d</i> | <i>s, z</i> | <i>s, z</i> |
| <i>*ćw, ǰw</i> | <i>s (θ), z</i> | <i>sp, zb</i> | <i>ś, ž</i> |

To the southwestern group belong the Persian languages and the other languages of southwestern and southern Iran (Bakhtiari and Lori, Fars dialects, Larestani, and Bashkardi). To the northeastern group belong Khotanese and Wakhi. All the others belong to the central group. Examples: Indo-Iran. **daća* '10,' OPers. **daθa*, Av. *dasa*, Khot. *dasau*; Indo-Iran. **aǰ^h am* 'I,' OPers. *adam*, Av. *azəm*, Khot. *aysu*; Indo-Iran. **aćwa*- 'horse,' OPers. *asa*-, Median, Av. *aspa*-, Khot.

aśśa-, Wakhi *yīš*; Indo-Iran. **-ijwā(n)*- 'tongue,' OPers. *hizān*-, MPers. *izbān*, Khot. *biśāa*- /βiśāa-*l*.

Characteristic Features of Morphology and Points of Syntax

Avestan and Old Persian are still of the Indo-European type (like Greek, Old Indic), with complex morphologies.

In the nominal and pronominal systems, Old and Young Avestan have three numbers (singular, dual, plural), eight cases (nominative, accusative, genitive, dative, instrumental, ablative, locative, vocative); in Old Avestan, the ablative singular has a distinct ending only in the *a*-declension (*-āt*), while in Young Avestan, the final *-t* is found in all declensions. Old Persian has only six cases and was also losing dual forms.

The Avestan verbal system is based on three stems: present, aorist, and perfect (e.g., *kar*- 'to do': PRES *kərənao*-, AOR *kar*-/ *icar*-, PERF *ca-kar*-). All the Indo-Iranian moods are preserved. There is a past participle in *-ta*- (e.g., *kərəta*-) and several infinitives. In Old Avestan, the three stems are associated with different aspects: not completed event (present, imperfect, injunctive); completed event (aorist); present result of past event (perfect). In Young Avestan and Old Persian, the aspect system survives mainly in modal forms; in Old Persian, aorist modal forms form a suppletive paradigm with present forms. The Young Avestan present injunctive (present stem with secondary endings) and the Old Persian imperfect (augment plus present stem with secondary endings) became general narrative tense, expressing continuous or punctual actions and events, as well as anteriority, depending on context (YAv. *kərənaom* [-*aom* < -*aw-am*] 'do-1ST.SING,' OPers. *a-kunav-am* 'PAST-DO-1ST.SING' = 'I did, (when ...) I had done.')

An innovation common to Young Avestan and Old Persian is the optative used (with or without augment) to express habitual actions or events in the past (YAv. *apataiēn* < *a-pat-aiy-ant* 'PAST-fall-OPT-3RD.PL' = '(the demons) used to run about'; OPers. *avājanyā* < **ava-a-jan-yā-t* 'down-PAST-strike-OPT-3RD.SING' = 'he used to kill'). This usage continues in several Middle Iranian languages.

Old Persian is in the process of developing a split-ergative verbal system, with an ergative perfect tense contrasting with the ancient imperfect (*a-kunav-am* 'PAST-DO-1ST.SING' = 'I did/had done' versus *manā kar-t-am* 'I.GEN-DAT DO-PAST.PART-NOM.SING. NEUT' = 'I have done [it]').

The Middle and New Iranian languages exhibit a variety of developments that represent smaller or greater innovations compared with Old Iranian.

Points of Historical Phonology

We may distinguish between languages with conservative phonologies, characterized by the survival of surd stops and affricates after vowels (comparable to Italian in this respect), and languages where these are voiced or dropped (comparable to Spanish and French). Sogdian and Balochi are of the former type, the others at various stages of the latter.

Final syllables are lost to various degrees, e.g., OIran. NOM.SING **aćw-ah* ‘horse’: Av. *asp-ō*, OPers. *as-a*, Khot. *asś-ā*, Sogd. *asp-ī*, Parth., MPers., NPers. *asp*, Pashto *ās*.

The Indo-Iranian syllabic *r* <ṛ> developed differently from *ar* in most languages (Av. <arə>, OPers. <ar>, MPers. *ir/er*, *ur*, etc.; e.g., Av. *zərədaia-* ‘heart,’ Parth. *zirō*, MPers. *dil* [-rd > -l]; Av. *mərə-ta-* ‘dead,’ OPers. *mar-ta-*, MPers. *murd*).

All the Old Iranian diphthongs were monophthongized. Fronting and backing of vowels (umlauting) is common in both diachrony and synchrony.

In Chorasmian, Khotanese, and many modern East Iranian languages, the old palatal affricates *č* [tʃ] and *ǰ* [dʒ] became dental affricates *c* [ts] and *j* [dz], while new palatal affricates developed from the velar stops by palatalization.

The groups *xt*, *ft* became *γt*, *βt* (with various further developments) in Ossetic and eastern languages (Av. NOM.SING *duxt-a* ‘daughter,’ Sogd., Chor. *duyd-ā*, Khot. *dūt-a* /dūd-a/, Bactr. <logd-a> = *lugd-a*, Pashto *-lā* [< **luw* in *tər-lā* ‘uncle/aunt’s daughter’]; MPers. *haft* ‘7,’ Sogd. *aft*, Khot. *hauda*, Pashto *owá*, Shughni *ūvd*, Wakhi *ib*).

Some modern East Iranian languages also have retroflex series, e.g., sibilants *š* [ʃ], *ṣ* [ʂ], *ž* [ʒ], *ṣ̌* [ǯ]; affricates *ts*, *tš*, *tṣ*, *dz*, *dž*, *dṣ*, etc. In Pashto, the phonetic realization of the retroflex sibilants defines four dialect areas, going from southwest to northeast: *pašto*, *pašto*, *pašto*, *paxto*.

The development of consonant groups is especially varied:

θr: OPers. *ç*, MPers., etc., *s*; Sogd. (and several mod. dialects) *š*; Khot. *dr-*, *r-*; Parth., Bactr. *hr*; Ossetic *rt* (e.g., OIran. NOM.SING **puθr-ah* ‘son’: OPers. *puç-a*, MPers. *pus*, NPers. *pesar* [changed after *pedar* ‘father’]; Sogd. *pāš-ī*, Khot. *pār-ā*, Parth. *puhr*, Bactr. <pouro> = *puhr*; Oss. *firt*; OIran. **θrājah* ‘three’: MPers. *sē*, NPers. *se*; Sogd. *šē*; Khot. *drai*; Parth. *h(e)ray*, Bactr. <uarēio> = *harēy*, Shughni *aray*; Oss. *ærtæ*, Wakhi *truy*; Munji *ṣiray*).

rs: remains in Ossetic and several eastern languages but commonly becomes *š* in western languages; Pashto and some others have *št* (e.g., **pṛsa-* ‘ask,’ Pashto *pušt-*); the Shughni group has variously *šc*, *ws* (Shughni *pešc*, Roshani *paws*).

rt: frequently remains or becomes *rd* (Wakhi *mərt* ‘dead’, MPers. *murd*); it becomes retroflex *t* [t̪] in several eastern languages (e.g., Khot. *muḍa-*), which further becomes retroflex flap *r* [ɾ] (e.g., Pashto *mṛə*), or loses retroflexion and becomes *d* or *g* (Shughni *mūd*, Roshani *mewg*).

št: Oss. *st*; Bactr., Pashto *t*; Shughni group *št*; Yidgha-Munji *škʷ*, *šč* (Av. *ašta* ‘eight,’ OPers. *aštā*, MPers., NPers. *hašt* [with *h-* from *haft* ‘seven’], Bactr. <atao> = *ata*, Pashto *atā*, Shughni *wašt*, Yidgha *aščo*).

Points of Historical Morphology and Syntax

Three genders are still found in Sogdian and Khotanese. Two genders are found in Bactrian and Chorasmian and in many modern languages. Persian and many modern languages have no gender.

The dual remains in Sogdian and Chorasmian. The numerative forms used only after numerals in Sogdian and Pashto incorporate old dual forms.

In several Middle and New Iranian languages, marked plural forms are restricted to animate nouns or are used to express individuality.

Definite articles from demonstrative (or relative) pronouns are found in Sogdian, Bactrian, Chorasmian, and Ossetic. In the modern languages, indefiniteness is usually expressed by a suffixed ‘one.’ Definiteness (topicalization, etc.) is expressed in a variety of ways (which is addressed later, in the section on direct object marking).

The western languages (Middle Persian, Parthian, Bactrian) have reduced the older, six-case system to a two-case system (direct case for the older nominative and accusative[?] and oblique for the other cases), while the eastern languages (Tumshuqese, Khotanese, Sogdian) preserve the Old Persian-type six-case system. In several declensions in Sogdian, in Chorasmian, and in the later stages of Khotanese, the nominative and accusative singular are no longer distinguished (exceptions in the pronouns). Several modern languages preserve the two-case system, sometimes with the addition of one or more local cases. Balochi and other eastern languages have nominative, genitive (possessive), oblique, and a case for the direct/indirect object. Ossetic, to some extent influenced by Caucasian languages, has the largest number of cases, including nominative, genitive, and dative (grammatical cases) and (local cases) allative, ablative, inessive, adessive, equative (expressing language and likeness), and comitative (only Iron).

In Sogdian, a system of light versus heavy stems developed: stems containing at least one long vowel or a diphthong (including *ar*, *an*, *am*) attracted the

stress, causing loss of short final vowels (e.g., NOM.SING *asp-i* 'horse,' LOC.SING *asp-yá*, versus NOM.SING *mēθ* 'day,' LOC.SING *mēθ-i*; IMPERF.3RD.SING *wan-á* 'did' versus *wēn* 'saw').

Several Middle and New Iranian languages use affixes to mark the direct object, depending on degree of specificity and definiteness. The most common contrast is unmarked = indefinite versus marked = definite. The markers are of three main types, all older indirect object markers: (1) ending derived from older case ending (usually *-e*, *-i*); (2) from older prepositions meaning 'to' (Manichean Parthian [less commonly Middle Persian] *ō* and Bactrian *abo* from OIran. *abi* 'to([ward])'; Bashkardi *be-* from OIran. *pati* 'to([ward])'); (3) from older prepositions meaning 'from, on account of,' etc. NPers., Balochi *-rā* [from OPers. *rādiy* 'on account of,' MPers. *rāy*, indirect object marker]; Shughni group *a(z)-* [OIran. *hača* 'from']).

In the eastern Middle Iranian languages, as well as in several modern East Iranian languages, the noun has two distinct declensions, one going back to the old vocalic stems, the other to extended *ka*-stems (feminine variously, *kā-*, *kī-*, *čī-*, and, analogically, *čā*-stems).

The western languages (Middle Persian, Parthian, Bactrian) continue, to varying extents, the Old Persian use of the relative pronoun as relative particle or *ezafe* (Mid. Pers. *ī*, Parth. *čē*, Bactr. *i-*). In New Persian and related dialects, the use of the *ezafe* is common. The so-called inverted *ezafe* found in some languages is the oblique case, e.g., Mazanderani *pēr-e kiya* 'father-OBL house' = 'the father's house.' Possession is of the type 'I have' or 'for me is.'

In verb morphology, innovations include the development of new marked present (continuous) tenses and the restriction of the unmarked old present to, sometimes remarked, modal functions; restructuring of the stem systems into pairs of intransitive-passive and transitive-causative verbs; the restructuring of the past tense systems by the addition of the new perfect system seen in Old Persian; and, when this becomes the regular past narrative tense, the creation of a new perfect system, using various strategies.

Continuous or progressive tenses are marked already in Sogdian, where a particle meaning 'being' is added to the personal forms (*ḍār-am-skun* 'hold.PRES-1ST.SING-PROG' = 'I am holding'). In Khotanese, petrified participles meaning 'sitting,' 'standing,' 'going' probably modify the verb in the same sense. In the modern languages, a variety of affixes are found. In Persian and the southwestern dialects, we find prefixes originally meaning 'always' (Class. NPers. *hamē* and *mē-*, NPers. *mī-*; Bakhtiari and Lori *ī-* and *ei-*). The Tati and Caspian dialects use

a present participle in **-ende* + copula (e.g., **kun-ende-i* > *kennī* 'do-PRES.PART-COP.2ND.SING' = 'you are doing, you do'). Several languages use constructions of the type 'be in doing': Gilaki *kār-a āmon dar-a* 'work-CONN come.INF in-COP.3RD.SING' = 'he is coming'; Larestani *a-kerdā-em* 'a-doing I am, I am doing'; NBashk. *a-kerdén-om*, SBashk. *be-kért(én)-in*; Balochi *raw-ag-ā int* 'a-go-ing he is, he is going'. Some eastern languages use a suffix derived from 'stand' (Yaghnobi *-išt*, e.g., *šāw-om-išt* 'go-1ST.SING-CONT' = 'I am going,' imperf. *a-šāw-i-m-išt* 'PAST-go-IMPERF-1ST.SING-CONT' = 'I was going').

The old present acquires modal functions and can be marked (e.g., NPers. subjunctive *kon-am* 'do-1ST.SING' or *be-kon-am* 'SUBJ.do-1ST.SING' = 'I may do').

The imperfect is still characterized by the augment in Middle Persian (only example *a-ger-īy* 'PAST.do-PASS.3RD.SING' = 'was made,' cf. OPers. *a-kar-īya*), Chorasmian, Sogdian, and Tumshuqese (*a-cchu* 'I went'), and in modern Yaghnobi (*a-kún-i-m* 'PAST-do-IMPERF-1ST.SING' = 'I did'). In Chorasmian and Sogdian, the augment is lost in verbs without original prefix; in verbs with prefix, the augment appears as lengthening of the vowel of the original prefix (e.g., Sogd. *θbar-*, IMPERF *θ-ā-bar-*, Chor. *haβir-*, IMPERF *h-ā-βir-* 'gave,' cf. OPers. *frā-bara* < *fra-a-bara* 'forth-PAST-carry-3RD.SING' = 'he gave') or as *m-* (e.g., Sogd. *m-anγaz-* 'began', Chor. *m-ikk-* 'did', analogically from forms with prefix *ham-*, cf. OPers. *ham-a-taxša-īy* 'PREV-PAST-labor-1ST.SING.MID' = 'I labored'). Several modern languages and dialects have imperfects with a suffix *-i-*, which may continue the Old Iranian preterital optative (e.g., Yaghnobi IMPERF *a-šāw-i-m* 'I was going,' PRES.SUBJ *šāw-om* '(that) I go').

In Middle Persian, Parthian, and Bactrian, the (narrative) past tense system (completed action) is based on the older split ergative seen emerging in Old Persian; Sogdian and Choresmian instead use the auxiliary 'to have' in transitive constructions and 'to be' in intransitive ones (Sogd. *ək-t-u-ḍār-am* 'do.PAST-ACC. NEUT.SING-have.PRES-1ST.SING' = 'I did, I have done'); and Khotanese has a form with an originally active (possessive) participle plus 'to be' (*dā-t-aimā* 'see-PAST-1ST.SING.MASC' = 'I saw, I have seen' < **dita* 'āh ahmi' 'having-seen I am'). As the old original perfect replaced the inherited imperfect (injunctive) as past narrative tense, new strategies were invented to express the perfect. The most common is the use of a past stem extended by *-ag* > *-e*, etc. (NPers. *kārd-am* 'I did', *kard-é-am* 'I have done') or *-ss-* (from the older 'stand'; Larestani *če-d-e* 'go-PAST-PERF.3RD.SING' = 'he has gone,' but *če-ss-em* [-*d-s* > *-ss-*] 'go-PAST-PERF-1ST.SING' = 'I have gone,' Kumzari *zur-s-e* [*zur-* < *zat-*] 'he has struck').

Punctual aspect is often marked by the prefix *b-* or similar, which also marks modal functions in many modern Iranian languages.

The old modal forms were generally preserved in Middle Iranian: subjunctive, optative, and imperative everywhere; the injunctive (with various modal functions) in Chorasmian, Sogdian, and Khotanese. In the modern languages, old modal forms survive as archaisms (NPers. *zende b-ād* 'alive be.3RD.SING. SUBJ' = 'long live!'), and, instead, modal functions are left unmarked in contrast to the marked present (continuous) or are marked by a prefix, often *b-*.

Directional prefixes are common, more or less closely connected with the verb. For instance, the Late Khotanese adverbs *vā*, *ttā*, *hā* and the Pashto prefixes *rā*, *dār*, *wār* indicate the direction of the action or movement toward first, second, or third person and can also simply substitute for the personal

pronouns. Ossetic has a complex system of spatial prefixes indicating the direction of the motion relative to location and speaker.

Bibliography

- Brill (1958). *Handbuch der Orientalistik. I. Iranistik. IV. 1. Linguistik*. Leiden and Cologne: Brill.
- Schmitt R (ed.) (1989). *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert.
- Schmitt R (2000). *Die iranischen Sprachen in Geschichte und Gegenwart*. Wiesbaden: Reichert.
- Skjærvø P O (2005). 'Iranian languages and their development.' In Yarshater (ed.) vol. 7.
- Windfuhr G (ed.) (2005). *The Iranian languages and typology*. London: Curzon.
- Yarshater E (ed.) (1982–). *Encyclopaedia Iranica*. London: Routledge and Kegan Paul.

Irish See: Goidelic Languages.

Iroquoian Languages

M Mithun, University of California,
Santa Barbara, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

The Iroquoian Family

The Iroquoian languages are indigenous to south-eastern and northeastern North America. The family consists of two major branches: Southern Iroquoian and Northern Iroquoian.

Southern Iroquoian is represented by just one language, Cherokee, spoken today primarily in Oklahoma and western North Carolina. There are clear dialect differences between western and eastern Cherokee and within each, many of which predate the forced march of the Cherokee from North Carolina to Oklahoma in 1838.

Northern Iroquoian consists of several subgroups. The first to break away from the main branch developed into Tuscarora, Nottoway, and Meherrin. The Tuscarora were first encountered in eastern North Carolina, but early in the 18th century most moved north to rejoin other Northern Iroquoians. Their descendants now live primarily in two communities,

one at Six Nations on the Grand River in Ontario, the other near Niagara Falls in New York State. Nottoway and Meherrin were once spoken near the Virginia and North Carolina coasts. All that remains of Nottoway are two wordlists recorded during the early 19th century. The only record of Meherrin is two town names, sufficient to identify the language as Iroquoian. The Meherrin merged with those Tuscarora who did not migrate north in the 18th century.

The next group to separate from Northern Iroquoian became the Huron. They comprised a confederacy of four nations totaling around 20 000 people when they encountered Champlain in 1615 in present southern Ontario. In 1649, they were decimated by the Five Nations Iroquois. Some survivors formed a settlement at Lorette near modern Quebec City. Others joined the remnants of neighboring Iroquoian nations, the Tionontati (Petun), Erie, and Neutral, and migrated west toward Detroit and ultimately into Oklahoma. This group became known as the Wyandot. Both Huron proper at Lorette and Wyandot in Oklahoma were last spoken in the 20th century.

The remaining Northern Iroquoians separated into several subgroups, whose territories extended across present New York State. These were the Five Nations (with Tuscarora, the Six Nations), members of the League of the Iroquois. To the west were the Seneca and Cayuga. In the center were the Onondaga, near modern Syracuse. To the east were the Oneida and the Mohawk. The languages are now mutually unintelligible, though they share many structural features. Seneca is now spoken in three communities in western New York: Cattaraugus, Allegany, and Tonawanda. Most Cayuga left New York State after the Revolutionary War, some going to Six Nations in Ontario, and others to Oklahoma. The language is now spoken primarily at Six Nations. It was last spoken in Oklahoma late in the 20th century. Onondaga is spoken at Six Nations and at Onondaga south of Syracuse. Oneida is spoken near London, Ontario, and Green Bay, Wisconsin. There are six Mohawk areas: Kahnawà:ke and Kanehsatà:ke in Quebec; Ahkwesáhsne, which straddles Quebec, Ontario, and New York State; and Ohswé:ken (Six Nations), Thaïentané:ken, and Wáhta' in Ontario. Most speakers now live in the first three.

Several other Iroquois languages are known to have existed as well. In 1534, Jacques Cartier encountered people along the St Lawrence River around present Quebec City. Vocabulary in his ship's logs and appended wordlists indicate that these people, now known as the Laurentian, spoke several Northern Iroquoian languages, at least one of which was not ancestral to any of the modern languages. When Champlain returned to the area in 1603, these people had disappeared. They are, however, the source of the name Canada. Another group, the Susquehannock or Andaste, were encountered by Captain John Smith about 1615 in the lower Susquehanna Valley in Pennsylvania. Their language, known through a wordlist recorded by the Swedish missionary Johan Campanius in 1696, was last spoken in the mid-18th century.

Phonology

The consonant inventories of the languages generally consist of one series of obstruents (reflexes of *t, *k, *tʃ, *kʷ, *s), one of resonants (*n, *r, *w, *j), and one of laryngeals (*h, *ʔ). Voicing is not distinctive. Noteworthy is the lack of labials. Most of the languages have four oral vowels (*i, *e, *a, *o) and two nasal vowels (*ɛ, *ɔ). In Proto-Northern-Iroquoian, stress was penultimate, and open, stressed syllables were lengthened.

There have been several innovations of interest. One is stress in Cayuga. Cayuga stress placement

depends on syllable count from both edges of the word. Primary stress falls on the penultimate syllable, providing that it is even-numbered counting from the beginning of the word: *satkáhtoh* 'Look!' If the penult is odd-numbered, it is stressed only if it is open: *senhotǫ:koh* 'Open the door!' If the penult is odd-numbered and closed, stress is antepenultimate: *sasaky'atáwihsih* 'Take your shirt off!' Epenthetic vowels, added after the establishment of basic penultimate stress in Proto-Northern Iroquoian, are not counted, like the *-a-* stem joiner in *sakyá'tawi't* 'Put your shirt on!' Stress on the second of two adjacent vowels moves to the first: *sasanáháowe:k* 'Put your hat back on!' If conditions are not met for penultimate stress, and there is no antepenultimate syllable, the word carries no stress: *sahsyǫ* 'you returned.'

Cayuga shows another interesting innovation. In odd-numbered syllables closed by a laryngeal, the laryngeal feature spreads leftward over the entire syllable. If the laryngeal is h, the full syllable is devoiced. (Devoicing is shown orthographically by underlining.) If the laryngeal is glottal stop, the full syllable carries creaky voice, indicated here by a wavy underline: *wahsí'tá'keh* 'on its foot.'

Several of the languages have developed distinctive tone under the effect of laryngeals. In Mohawk, stressed syllables generally show high pitch (if short) or rising pitch (if long). Stressed syllables closed by a laryngeal (glottal stop, or h followed by a resonant), however, developed a special pitch contour: a rise followed by a deep fall. The triggering laryngeal subsequently disappeared before a consonant. There are now contrasts, such as *oká:ra* 'story' (with rising tone) and *okà:ra* 'eye' (**okáhra*'). (Mohawk examples are given here in the community orthography. Nasalized vowels ʌ and ʊ are written en and on respectively, glottal stop with an apostrophe ', vowel length with a colon :, stress accompanied by high or rising tone with an acute accent, and stress accompanied by falling tone with a grave accent. The palatal glide j is represented by i. Other symbols have approximate IPA values.) Western Cherokee underwent more complex changes resulting in tone. Feeling (1975) distinguishes three level tones (2, 3, 4), a rising tone (23), and two falling tones (32, 21, written 1), as in *a¹hyv²³ki³di²ya* 'he's capturing him.'

Morphology and Syntax

All of the languages are polysynthetic. Verbs in particular can be composed of large numbers of meaningful parts (morphemes), like Mohawk *enhske-rhar-átst-en-* 'you will promise me' (FUT-2. sing/1.sing-expect-CAUS-BEN.APPL-PER). All verbs contain at least three parts: a pronominal prefix, a verb root,

and an aspect suffix. The pronominal prefix refers to the core arguments of the clause: *-hske* ‘you/me’ in the verb above. The prefix appears in the verb whether or not there are coreferential independent nominals in the clause as well, but it is fully referential in its own right.

There is little evidence of Subject or Object categories. The pronominal prefixes show grammatical Agent/Patient patterning. The semantic basis underlying the system is still clear. Agent prefixes typically represent participants who are actively in control and instigating events or states, as in Mohawk *ra-tahónhsatats* ‘he is listening,’ *ra-ná:ie* ‘he is conceited,’ or *rá-hsere’s* ‘he is chasing it.’ Patient prefixes typically represent participants who are affected by a situation but not in control: *ro-thón:te* ‘he hears it,’ *ró-ta’s* ‘he is sleeping,’ or *ró-hsere’s* ‘it is chasing him.’ The system is now fully routinized, however. Speakers do not judge degrees of control as they speak; they simply learn which set of prefixes to use with each verb.

Agent prefixes are used in both events and states (*jump*, *be conceited*), and Patient prefixes are used in both events (*holler*) and states (*be poor*). The system is thus basically Agent/Patient rather than Active/Stative. Superimposed on it, however, is one element of Active/Stative patterning. Some verbs occur only in the Stative: *rahseró:hen* ‘he is quick-tempered,’ *ró:ten* ‘he is poor.’ Others occur in all three aspects: Habitual *ratè:kwahs* ‘he escapes,’ Perfective *wahatè:ko* ‘he escaped,’ Stative *rotè:kwen* ‘he has escaped.’ With verbs like *escape*, which describe a change of state, the Stative aspect form typically has a Perfect meaning: *he has escaped*. All Perfect Stative verbs occur with Patient prefixes, whether their Habitual and Perfective forms appear with Agents or Patients.

The Northern Iroquoian languages show extensive, productive noun incorporation. The Southern language Cherokee shows traces of incorporation embedded in the lexicon, indicating that incorporation was productive in Proto-Iroquoian. Incorporation is a process whereby a noun stem is compounded with a verb stem to form a larger verb stem: *-itsker-onti-* ‘saliva-throw = to spit.’ There is no explicit specification of the semantic role of the incorporated noun; it simply indicates the involvement of a type of entity, often as a semantic patient, but sometimes as an instrument, location, source, or goal. Examples can be seen in Mohawk *wa-ho-n-itsker-ón:ti-* FACTUAL-MASC.SING.AGT-MIDDLE-saliva-throw-PER ‘he saliva-threw = he spit’ and *ka-hseriie’t-áner-en* NEUTER-cord-tie-STATIVE ‘it is cord-tied.’ Incorporation is used pervasively to create new vocabulary and also to manipulate the flow of information through discourse. Important new participants are typically introduced with independent nominals,

but those that are already part of the scene or of peripheral importance may be carried along as incorporated nouns.

Lexical categories are clearly distinguished by their internal morphological structure. Particles, by definition, are morphologically unanalyzable (*tóka* ‘maybe,’ *oh* ‘what,’) though they may be compounded (*nek tsi* ‘the only’ for ‘because’). Nouns contain a prefix specifying the gender of the referent or its possessor, a noun stem, and a noun suffix: *ka-nà:tsi-a* NEUTER-kettle-NOUN.SUFFIX ‘kettle,’ *ake-nà:tsi-a* 1.SING.ALIENABLE.POSS-kettle-NOUN.SUFFIX ‘my kettle.’ Alienable and inalienable possession are distinguished. Verbs follow an entirely different pattern, and can be quite complex morphologically. In addition to the obligatory pronominal prefix, verb root, and aspect suffix, they may contain various prepronominal prefixes. In the Northern languages, these are the Partitive, Contrastive, Coincident, Negative, Translocative, Factual, Duplicative, Future, Optative, Cislocative, and Repetitive. They may contain a Middle, Reflexive, or Reciprocal prefix. As seen above, they may contain an incorporated noun stem. Following the verb root, there may be one or more derivational suffixes, such as an Inchoative, Reversive, Causative, Instrumental Applicative, Benefactive Applicative, Distributive, or Purposive. After the aspect suffix, there may be a postaspectual suffix: a Past, Continuative, or Progressive. Such structure can be seen in the Mohawk *a-khe-’nikonhr-áks-a’t-e* OPT-1.SING/INDEF-mind-be.bad-CAUS-PER ‘I would insult someone,’ or *ia’-te-iako-hah-a-hiia’k-on-hátie* TRANS-LOCATIVE-DUPLICATIVE-FEM.PAT-road-STEM JOINER-CROSS-STATIVE-PROG ‘she was crossing the street.’

The morphological composition of particles, nouns, and verbs is thus entirely distinct. Noun stems never appear in the verb stem position of verbs, and verb stems never appear in the noun stem position of nouns. The match between internal morphological form and external syntactic function, however, is not isomorphic. Some morphological particles function syntactically and semantically as nominals, such as *è:rhar* ‘dog.’ Morphological verbs can function syntactically as predicates, as nominals (without further derivation), or as full clauses, as below.

| | |
|-----------------------|---------------------------------|
| Nahón:ne’ | tehniíáhse |
| n-a-honn-e-’ | te-hni-iahse |
| PART-OPT-MASC.PL.AGT- | DUPLICATIVE-MASC.DU. |
| go-PER | AGT-be.together.STAT |
| ‘they would go there’ | ‘they two (males) are together’ |
| niristi:sere’s | nahshakotihahónnien’. |
| ni-ris-ti’ser-e’-s | ne = a-hshakoti-hah-onni-en-’ |

MASC.DU.AGT-steel- the = OPT-MASC.PL/3.DPL-road-
 drag-STAT-DIST make-BEN.APPL-PER
 they (males) steel the they would road make for
 drag around them
 'They would go serve as guides for two surveyors.'

Finally, there is no basic, syntactically defined word order. In part because of the richness of the verbal morphology, the proportion of verbs to nouns and of predicates to nominals is much higher in Iroquoian languages than in many languages of Europe and Asia. There are few oblique or adjunct nominals. When clauses do contain multiple constituents, all orders are not only possible, they can also all be seen to be pragmatically motivated by the discourse at hand.

Bibliography

- Chafe W (1967). Seneca morphology and dictionary. *Smithsonian Contributions to Anthropology* 4. Washington, D.C.
 Feeling D (1975). *Cherokee-English dictionary*. Tahlequah, OK: Cherokee Nation of Oklahoma.

- Foster M (1974). *National Museum of Man Mercury Series 20: From the earth to beyond the sky: an ethnographic approach to four longhouse Iroquois speech events*. (Cayuga). Ottawa: Canadian Ethnology Service.
 Lounsbury F (1953). *Oneida verb morphology*. New Haven: Yale University Press.
 Michelson K & Doxtator M (2002). *Oneida-English and English-Oneida dictionary*. Toronto: University of Toronto Press.
 Mithun M (1979). 'Iroquoian.' In Campbell L & Mithun M (eds.) *The languages of Native America*. Austin, TX: University of Texas. 133–212.
 Mithun M (1999). 'Iroquoian family.' *The languages of Native North America*. Cambridge: Cambridge University Press. 418–430.
 Mithun M & Henry R (1982). *A Cayuga grammar*. Brantford, ON: Woodland Indian Cultural and Educational Centre.
 Rudes B (1999). *Tuscarora-English English-Tuscarora dictionary*. Toronto: University of Toronto Press.
 Woodbury H (2003). *Onondaga-English English-Onondaga dictionary*. Toronto: University of Toronto Press.

Italian

A L Lepschy and G Lepschy, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Italian is the official language of the Republic of Italy. It is spoken in Italy (including the Republic of San Marino and the Vatican City), and outside Italy, in the Canton Ticino (it is one of the official languages of the Swiss Confederation) and, with different degrees of vitality, in areas such as those of Nizza, the Principality of Monaco, Corsica, Istrian and Dalmatian towns, and Malta. It also survives in Eritrea, and where there are large communities of Italian emigrants, particularly in the United States (about four million speakers), and in Canada, Argentina, Brazil, and Australia (about half a million each).

Number of Speakers

The inhabitants of the Italian Republic numbered about 57 million in 2002. The number of speakers of Italian is, however, more difficult to establish, mainly because there is no simple way of relating the use of the Italian language with Italian ethnic origin or cultural allegiance, in the case of emigrants,

and even with Italian citizenship in the case of the residents of Italy. For the latter, the problem is posed not so much by the linguistic minorities within the boundaries of the Italian republic (speakers of German, ca. 280 000, mainly in South Tyrol; of Occitan and Franco-Provençal, ca. 115 000; of Slovene, ca. 53 000; of Serbo-Croatian, ca. 3000; of Albanian, ca. 100 000; of Greek, ca. 30 000; of Catalan, ca. 15 000), but rather by the presence of the so-called 'Italian dialects.'

Genetic Relationships

The term 'dialect,' in the Italian tradition, is used to refer not to different varieties of the same language but, rather, to 'siblings' of Italian. Italian is based on the literary Tuscan (more specifically Florentine) of the fourteenth century. This in turn derives from Spoken Latin and is therefore a Romance language. But Latin during the period from the sixth to the ninth century gave origin, in Italy, to a myriad of Romance languages, which can be classified into over 15 major groups (broadly coinciding with the Italian regions) such as Piedmontese, Lombard, Venetian, Ligurian, Emilian, and so on. These Italian 'vernaculars', in Italian 'volgari' (a term used to designate the living

languages, in the Middle Ages, as opposed to Latin, which was no longer native to any group of speakers, but was the standard written language) in their modern forms constitute the Italian dialects (Sardinian and Friulian are sometimes classified as Italian dialects, sometimes as separate Romance languages). The differences between these dialects can be very pronounced. Speakers of Bolognese and of Neapolitan find each other's dialects as unintelligible as would, say, speakers of French and Spanish.

History

External History

Literary Florentine, as used by the great Tuscan writers of the fourteenth century, Dante, Petrarch, and Boccaccio, was gradually accepted as the national standard language and finally codified as such at the beginning of the sixteenth century. The '*Questione della lingua*' (question of the language) was a central concern for the culture of the time with ideological and political, as well as linguistic and literary, implications. Against the proposers of contemporary Tuscan, and of a supraregional '*lingua cortegiana*' (language of the courtiers), the solution that prevailed was the one defended by the Venetian humanist, Pietro Bembo. The model he upheld was archaic Tuscan, which did not imply subservience to any of the rival modern political powers in Italy and was consistent with the principles of classicism and imitation then prevailing.

This had important consequences for the linguistic history of Italy. On the one hand Italian appears to be exceptionally 'conservative': it must be almost unique, among modern national languages, in having changed so little through eight centuries of documented history, and in allowing a twentieth-century reader to approach thirteenth-century Tuscan texts with minimum effort. On the other hand (and this is perhaps the condition for such stability) this language remained a standard written model, accessible only through literacy, like Latin, rather than acquired as a native language. The vast majority of the people, who were illiterate, did not have access to the literary language, and in speech used exclusively their native dialect. It has been calculated that at the moment of political unification in 1860 the number of the inhabitants of the country who could use Italian was at most about 10%, perhaps as low as 2.5%. In 2005, after 145 years of existence of a unified Italy, Italian was assumed to be known to all Italians, but statistics indicated that almost one-half of the population still preferred to use a dialect rather than the national language.

Internal History

The main points to be mentioned in the change from Latin to Italian include the following.

Phonology The opposition of long and short vowels in Latin was replaced by a distinction determined by stress and syllable structure: long vowels are obligatory in free stressed nonfinal syllables, short vowels in other conditions (i.e., in checked stressed, free stressed final, and all unstressed syllables). In stressed syllables the five Latin vowel qualities became seven, with long \bar{e} and \bar{o} giving midhigh [e] and [o], short \breve{e} and \breve{o} giving midlow [ɛ] and [ɔ] (these broke into [jɛ] and [wɔ] in free syllables). In unstressed syllables only five vowels are used: [i], [e], [a], [o], [u]. The consonant system undergoes the following main changes: assimilation (e.g., [kt] > [tt], as in *factum* > *fatto*); palatalization and assibilation before front vowels (as in *cenam* [ke:nam] > *cena* [tʃena]); *hodie* > *oggi*, *medium* > *mezzo*; sonorization, which applies unsystematically, as in *stratam* > *strada*, but *amatam* > *amata*. Initial *h* and final consonants (apart from nasals and liquids in proclitics: *un*, *per*, *il*, etc.) are dropped.

Morphology The case system of the Latin declension disappears and prepositions are used instead: *di* *uomo* replaces Latin *hominis*. The neuter gender disappears. The Latin verb pattern is basically preserved, with the introduction of 'analytical' or compound forms for the passive (*è amato* for *amatur*), the future (*amare ho*, whence the new synthetic *amerò*, for *amabo*), the perfect (*ho amato* for *amavi*, although this survives as *amai*), and so on.

Syntax The freedom of Latin word order is reduced, as syntactic function is signaled by linear position rather than case endings: for 'Paul saw Peter' the only normal and unequivocal structure in Italian is '*Paolo vide Pietro*,' whereas in Latin *Petrum Paulus vidit* would be equally clear in any of the six theoretically possible combinations of these three words.

Written Records

The first dated text in an Italian vernacular is found in the account of a court case in Campania in the year 960 A.D. (the *Placito capuano*). The original record in Latin has been preserved; it includes the statement repeated by some witnesses in the vernacular: *sao ko kelle terre per kelle fini que ki kontene trenta anni le possette parte sancti Benedicti* ('I know that those lands within those boundaries which here are contained, for 30 years the party of St Benedict

owned them'). Several, mostly short vernacular texts, are found subsequently until the thirteenth century, when 'Italian' literature proper begins, with religious compositions, the poetry of the Sicilian School, and finally Tuscan poetry.

The vernacular must have been used in speech for a long time before the document of 960. There are several texts that appear to represent an attempt to fix in writing a kind of vernacularized Latin by lowering it toward the spoken language, or raising the latter to conform to the conventions of written Latin. One of the best-known documents of this 'compromise' is a riddle apparently jotted down in Verona, at the beginning of the ninth century on a page of a prayer book; it refers to the act of writing, under the guise of describing the work of a farmer: '*se pareba boues alba pratalia araba & albo uersorio teneba & negro semen seminaba*' (a plausible rendering, among many which have been proposed: 'he was driving oxen, he was plowing white fields and he was holding a white plow, and he was sowing black seed'); some features appear more Latin than vernacular (e.g., *-b-* for *-v-* in the imperfects, final *-n* in *semen*, etc.); others are more vernacular (e.g., fall of final *-t* in the imperfects, *e* in *negro*, etc.). However, many traits present in dialects are attributed to the substratum of pre-Latin idioms spoken in Italy at the time of Romanization. For instance, a Neapolitan feature such as *-nn-* for *-nd-* belongs to Oscan and is found in the Latin graffiti of Pompeii, preserved under the ashes from the eruption of Vesuvius in 79 A.D. This, however, does not entitle us to suggest that Neapolitan was spoken in the first century A.D.

Writing System

The Italian writing system derives from the Latin one. It uses the Roman alphabet as it was adapted for the vernacular during the Middle Ages. Palatalization created some problems, which are solved in the Italian spelling conventions, by using *-* in front of *i* and *e* – *c*, and *g* for the palatals [tʃ] and [dʒ], and *ch* and *gh* for the velars [k] and [g]. But the system does not render all the oppositions of Tuscan phonology. In particular there are individual letters that correspond to contrasting sounds: *e* to [e] and [ɛ], *o* to [o] and [ɔ], *s* to [s] and [z], *z* to [ts] and [dz]; the last two alveolar affricates, which are always long intervocalically, are represented in spelling sometimes by *z* and sometimes by *zz*. These points leave a trace in the history of the language. As Italian was adopted as a national language in its written rather than spoken form, it became established, in different parts of Italy, with phonological counterparts, for these 'ambiguous'

letters, which may be different from the Tuscan ones. Considering that contemporary Tuscan does not constitute an undisputed standard, Tuscan pronunciations like *b[ɛ]ne*, *ca[s]a*, or *[ts]io* are not felt, in present-day Italian, to be more correct than Northern ones like *b[e]ne*, *ca[z]a*, or *[dz]io*.

Individual Characteristics

Among the Romance languages Italian would appear to be typologically more similar to Spanish than to French.

For syntax, one frequently noted feature is that it is a 'pro-drop' language (i.e., it does not need to express the pronominal subject of a verb and that, consistently, it can put the subject after the verb). The adjective has two positions: postnominal if it is restrictive, and prenominal otherwise.

Morphology is conservative within a traditional Indo-European pattern; adjectives, nouns and verbs are subdivided into classes, without apparent semantic justification. An adjective may behave like *ross-o* (MASC SG), *ross-a* (FEM SG), *ross-i* (MASC PL), *ross-e* (FEM PL); or like *verd-e* (MASC and FEM SG), *verd-i* (MASC and FEM PL); or like *par-i* (MASC and FEM SG and PL). A noun may behave like *la cas-a* (FEM SG) *le cas-e* (FEM PL); or *il poet-a* (MASC SG) *i poet-i* (MASC PL); or *il libr-o* (MASC SG), *i libr-i* (MASC PL), and so on. A verb may behave like *cant-are*, or *ved-ere*, or *dorm-ire*. Each conjugation has a vast array of different forms according to person (*canto*, *canti*, *canta*), number (*canta*, *cantano*), tense (*canto*, *canterò*, and, with aspectual distinctions, *cantava*, *cantò*), and mood (*lui canta*, *che lui canti*). There is a complex pattern of agreement for gender and number that involves nouns, articles, adjectives, and past participles: *è arrivata una ragazza alta*, *è arrivato un ragazzo alto*. A notable feature in Italian derivational morphology is the productivity of evaluative suffixes (often called 'alteration'): the following forms are based on *libro*: *librone*, *libretto*, *librino*, *libruccio*, *libraccio*, etc.

Phonology

The rhythm of Italian is syllable timed. There are seven vowels, very similar to the cardinal ones: /i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/. In unstressed syllables the opposition of midhigh and midlow vowels is neutralized, the quality of the other vowels remains distinct. There are two semiconsonants: /j/ and /w/, and 21 consonants: /p/, /b/, /t/, /d/, /k/, /g/, /ts/, /dz/, /tʃ/, /dʒ/, /f/, /v/, /s/, /z/, /ʃ/, /m/, /n/, /ɲ/, /l/, /ʎ/, /r/. Typologically uncommon is the systematic opposition of long to short consonants, which applies to all the items listed

apart from six: /z/ is always short, and /ts/, /dz/, /ʃ/, /p/, and /k/ are always long intervocalically.

Illustration

Here follows a sentence (1) quoted for illustration:

- (1) *A Venezia è più facile che si senta parlare il dialetto che l'italiano.*
 'In Venice it is easier to hear people speaking dialect than Italian.'

Phonemic transcription: /a vve'nettsja ɛ ppju f'fatfile ke ssi 'senta par'lare il dja'letto ke ll ita'ljano/. Note that this sentence is pronounced in Venice without 'syntactic doubling' (i.e., the lengthening of initial consonants in specified conditions) and with some of the variations mentioned above in the section, 'Writing System': /a ve'netsja ɛ pju 'fatfile ke si 'senta par'lare il dja'letto ke l ita'ljano/. In the local dialect this would be: /a ve'nesja ze pju 'fasie ke se 'seŋta par'lar el dja'eto ke l ita'ljɑŋ/.

Note the gender and number agreement between articles and nouns; the use of the subjunctive governed by *facile che*; the interesting construction with *si*, which can be interpreted as an impersonal (with *si* acting as the indefinite subject of *senta*: 'one hears'), or as a passive, with the infinitive clause (*parlare*) acting as sentential subject ('speaking is heard'). Also, the subject of the infinitive need not be specified: '*si sente parlare*' 'one hears [someone] speak.' There is no dummy subject for *è* ('it is'), and the sentential subject clause introduced by *che* comes after the verb. The first *che* acts as a complementizer, and the second *che* as a conjunction within the comparative structure (*più ... che*).

Sociolinguistic Points

As mentioned above, in Italy there are many linguistic enclaves in which 'foreign' languages are spoken, and in some cases their use is 'protected' by special legislation. For the majority of Italians the traditional situation was one of diglossia (with the local dialect used in speech, and literary Italian in writing. After political unification, and particularly as a consequence of far-reaching social changes, such as internal migration (mostly from the south to northern industrial conurbations) and the influence of the mass media, Italian has been widely adopted in speech as well. Regional differentiation is clearly marked in phonology, identifiable in lexis, and less clearly noticeable in grammar. A colloquial variety of the language has developed that has been called 'popular Italian'; this appears to be gaining acceptability, and

some of its features are penetrating into the standard written language (e.g., *gli* is now frequently found in writing for 'to them,' and sometimes even for 'to her,' as well as for the traditional 'to him'). The dialects have been diagnosed often as terminally ill and on the point of demise. In fact, they have proved remarkably resilient in ordinary usage, and sometimes they appear to be taken up by people (including the young) as a way of reasserting their own group identity and reacting against an alienating process of national equalization. There has also been a striking vitality in dialect poetry, often using very local, individual forms of the dialect, rather than a generalized, regional variety.

Descendants

Italian played an important part in the formation of the lingua franca used in the past in the Mediterranean. Since the Renaissance there have also been varieties of '*lingua zerga*' employed by vagrants. A level of 'slang' is generally thought not to be popular in Italian because its functions were filled by the dialects. A curious form of English slang, called 'parlyaree' or 'polari,' traditional among sailors, actors, and gays (and now surviving in the form of lexical relics) was based on Italian. Italian-based creoles seem to exist or to have existed in Eritrea, Argentina, and Brazil (Harris and Vincent, 1987: 20).

Languages Influenced

The influence of Italian has been mainly felt, in all the major European languages, at the level of high culture, in the lexis of music and the figurative arts. In the Renaissance, and for a long time subsequently, an acquaintance with Italian was thought to be part of the cultural equipment of educated people in most European countries.

History of Linguistic Investigation

More than for other European languages, linguistic awareness and a discussion of linguistic matters (the *Questione della lingua*) has always been very relevant for Italian intellectuals. During the Renaissance the description and codification of the language was an important part of national culture, and the *Vocabolario* (1612) of the Accademia della Crusca was the first of the great national dictionaries to be published. Less important was Italy's contribution to linguistic studies (including the historical investigation of the Italian language and dialects) during the nineteenth and twentieth centuries.

Bibliography

- Cortelazzo M, Marcato C, De Blasi N & Clivio G P (eds.) (2002). *I dialetti italiani. Storia struttura uso*. Turin: UTET.
- De Mauro T (1970). *Storia linguistica dell' Italia unita*. Bari: Laterza.
- Devoto G (1978). *The Languages of Italy*. Chicago, IL: University of Chicago Press.
- Harris M & Vincent N (eds.) (1987). *The Romance languages*. London: Croom Helm.
- Holtus G, Metzeltin M & Schmitt C (eds.) (1988). *Lexikon der Romanistischen Linguistik, Bd IV. Italienisch, Korsisch, Sardisch*. Tübingen: Niemeyer.

- Lepschy G (2002). *Mother tongues and other reflections on the Italian language*. Toronto: University of Toronto Press.
- Lepschy A L & Lepschy G (1991). *The Italian language today*. London: Routledge.
- Maiden M (1995). *A linguistic history of Italian*. London: Longman.
- Maiden M & Parry M (eds.) (1997). *The dialects of Italy*. London: Routledge.
- Migliorini B (1984). *The Italian language*. London: Faber and Faber.

Italian as a Fusional Language

C Iacobini, Università di Salerno, Salerno, Italy

© 2006 Elsevier Ltd. All rights reserved.

In a fusional language words are inflected (and derived) by using affixes whose boundaries are difficult to identify due to the tendency of affixes to fuse with one another and with the root. The numerous occurrences of allomorphy and the tendency of morphs to simultaneously encode several meanings result in the fact that there is no one-to-one correspondence between morphs and morphemes, thus making linear segmentation difficult.

Other terms are sometimes used instead of *fusional*: *flectional*, *inflectional*, or *inflecting*, *inflective*. The term *fusional* should be preferred to others referring to inflection because both agglutinating and polysynthetic languages can be highly inflectional.

Morphological Typology

The fusional type is one of the main morphological types. The classification of languages by morphological types is part of the standard terminology of linguistics, but it is also strongly criticized by the majority of typologists for three main reasons: (1) because the classification criteria are rather vague and difficult to apply in a consistent way; (2) because the morphological type is defined in terms of mutual favorability of properties rather than of implicational correlations, resulting in a low predictive power; and (3) because morphological typology has a holistic background.

The vagueness of the classification in morphological types is shown by the lack of consensus on both the number of types and the number of parameters identifying them (the three most-used parameters are (1) the ratio of morphs to word forms, (2) the number

of morphemes to morphs, and (3) the degree of word-internal modification of morphs (Greenberg, 1954).

Modern linguistics has disavowed the ideological prejudice dating back to the 19th century, according to which the fusional morphological type was regarded as being superior to other types. Just like these, the fusional type is a combination of functionally interconnected features, which – as a whole – form an ideal construct characterizing (the whole, or some aspects of) the morphology of languages. Languages are rarely pure types; they usually mix elements of different types. Assigning a language to a specific type depends on the preponderance of features considered significant (the quantification of such features is a difficult problem to solve from a practical point of view).

Despite criticism, the classification by morphological types is convenient and widely used to rapidly identify a number of features that tend to cooccur in the morphology of a language, and also to assess the extent to which a language moves away from such ideal constructs, both in a synchronic and in a diachronic perspective (some authors argue that languages tend to move toward a typological goal (Dressler, 1985).

The Fusional Type

The best-known attempt to establish a list of features that cooccur in morphological types is the one made by Skalička (1966). The features that tend to cluster in languages displaying fusional morphology can be listed as follows:

1. Words are formed by a root and (one or more) inflectional affixes, which are employed as a primary means of indicating the grammatical function of the words in the language. Agreement is widely employed.

2. High degree of modification of internal morpheme boundaries, with a consequently difficult linear segmentation.
3. Tendency to cumulate morphological meanings in a single affix (with consequent asymmetry between the semantic and formal organization of grammatical markers).
4. Word-class distinction is maximal. Inflection is rich, regarding both the number of inflectional classes and the extension of paradigms.
5. Stem suppletion; many cases of both homonymy and synonymy among affixes; clear distinction between inflectional and derivational affixes.
6. A slight correlation with syntax can be seen in the relatively free word order (but there are also fusional languages with a fairly fixed word order).

The fusional type is differentiated from the isolating type by the use of bound morphs and the clear-cut distinction between word classes; it is differentiated from the agglutinating type by the kind of juncture between morphemes and the nonbiunivocal correspondence between morphs and morphemes. In the synthetic vs. analytic distinction, the fusional type tends toward the synthetic end.

The fusional type is largely represented in Indo-European languages, especially the most conservative ones. Latin, Slavonic, and Romance languages are the ones that are most often mentioned as good representatives of this morphological type. The main features identifying the fusional type will be described hereafter using examples drawn from Italian.

Affixal Inflection and Agreement

Italian makes wide use of inflectional affixes (nouns, verbs, adjectives, pronouns, articles are usually inflected, whereas adverbs are invariable). Inflection is obtained by replacing the word ending.

The two following sentences show an example of agreement in the singular and in the plural:

*quell-a buon-a tort-**a** è finit-a, l-a avrei mangiat-a
volentieri*
'that good pie is finished, I would have eaten it with
pleasure'
*quell-e buon-e tort-**e** sono finit-e, l-e avrei mangiat-e
volentieri.*

All the underlined elements agree in number and gender with the feminine noun in bold, the auxiliary agrees only in number, the adverb is invariable; it should be noted that agreement also affects the pronoun and the participle in the second clause.

Cumulative Exponence in Adjectival Inflection

The main adjectival inflectional class has four endings

bell-o 'nice' (sing MASC)
bell-i (PL MASC)
bell-a (sing FEM)
bell-e (PL FEM)

Each affix codes two grammatical meanings at the same time (i.e., gender and number). Regarding adjectival inflection, the Italian language is more fusional than other Romance languages. In Spanish, for example, the gender and the number are expressed by two separate morphs, and the morph for the plural is the same for the two genders:

hermos-o (sing MASC)
hermos-o-s (PL MASC)
hermos-a (sing FEM)
hermos-a-s (PL FEM)

The other important inflectional class employs two affixes: *-e* for the singular, *-i* for the plural. Homonymy is therefore observed between the affixes from the two classes (*-e* can signify both PL FEM and sing MASC or FEM; *-i* can signify both PL MASC or FEM). Compared with Latin, Italian has lost the ability to form the comparative through affixation (the few affixal comparatives in use are instances of fused exponence: in *migliore* 'better' and *minore* 'minor' it is not possible to segment the lexical base from the comparative suffix), whereas the superlative form is productively obtained by using the suffix *-issimo* (*bello/bellissimo* 'very beautiful').

Noun Inflection and Affixal Homonymy

A recent classification of Italian nominal inflection proposed by D'Achille and Thornton (2003) – superseding the very unsatisfactory traditional classification based on the endings of the singular form of nouns – distinguishes five inflectional classes (defined as a set of lexemes whose members each select the same couple of endings for singular and plural), plus a sixth class consisting of invariable nouns (the bottom line of Table 1 shows the percentage of types of each class in the Italian basic vocabulary).

Class 1 consists of masculine nouns in the overwhelming majority (*librol-i*), with very few feminine exceptions (*manol-i*). Class 2 is made up of feminine nouns only. Class 3 consists of masculine nouns by approximately 45% (*fiorel-i*), of feminine nouns by the same percentage (*siepel-i*), and of ambigeneric by 10% (*cantantel-i* may be used both for a male or a female singer). Class 4 is mostly

Table 1 Nominal inflectional classes and percentage of types of each class in the Italian basic vocabulary

| Nominal Inflectional Class | 1 sing/PL | 2 sing/PL | 3 sing/PL | 4 sing/PL | 5 sing/PL | 6 sing/PL |
|----------------------------|-----------|-----------|-----------|-----------|-----------|------------|
| Endings | -o/-i | -a/-e | -e/-i | -a/-i | -o/-a | invariable |
| % | 41.2% | 30.3% | 20.6% | 1.2% | 0.2% | 5.4% |

Table 2 Low predictability of gender distinction on the basis of the singular ending of nouns

| | |
|-------|--|
| -o | prevailing MASC, but also FEM (<i>mano, foto</i>) |
| -a | prevailing FEM, but also MASC (<i>papa, clima</i>) |
| -e | equally divided between MASC (<i>fiore, cantante</i>) and FEM (<i>siepe, cantante</i>) |
| other | equally divided between MASC and FEM |

composed of masculine nouns (*poeta/-i*) but includes a couple of feminine nouns as well (*ala/-i*); in class 5 nouns are masculine in the singular and feminine in the plural (*uovo/-a*); all the nouns ending in a stressed vowel or in a consonant (class 6) are invariable, but class 6 also includes invariable nouns ending in each of the four vowels used as inflectional endings, except for /a/ in feminine nouns (MASC: *cinema, golpe, kiwi, stereo*; FEM: *consolle, ipotesi, radio*). Although there seems to be a tendency to express the singular by means of back vowels (*o* MASC, *a* FEM) and the plural by means of front vowels (*i* MASC, *e* FEM), the cases of affixal homonymy are quite numerous (remember that class 3 uses front vowels both for the singular and for the plural, and it includes masculine and feminine as well as ambigeneric nouns). Although on the one hand, the recent trend toward an increase of invariable nouns ending by a vowel curbs the ratio of inflected nouns, on the other it decreases the degree of correlation between form and inflectional class. The low predictability of both the inflectional class and gender is clearly shown in Table 2, which refers to singular (Thornton, 2001).

Traces of Case in Pronouns

Italian has lost a fusional feature that characterizes Latin nominal inflection: the case. However, one can detect traces of case in the Italian pronominal system. The choice of the correct form of certain pronouns (Table 3) demands a decision that depends on whether pronouns express a subject, a direct, or an indirect complement.

The system of stressed pronouns distinguishes three persons for the singular and three for the plural. The third-person singular has distinct forms on the basis not only of gender, but also of the feature \pm human. The third-person plural subject differentiates gender but not the \pm human feature.

In the spoken language, the pronouns *lui, lei, and loro* tend to broaden their functional scope and are frequently used both as subject and nonsubject, even for nonhuman referents.

The Table 4 shows clitic pronouns with the function of direct and indirect complement.

The use of such pronouns is complicated by rules of reciprocal ordering as well as by allomorphy: all clitic pronouns that end in *-i* replace the vowel with /e/ if followed by another clitic.

ti racconto una storia 'I'll tell you a story'

te la racconto 'I'll tell it to you'

te lo porto 'I'll bring it to you'

portaglielo 'bring it to him/her'

The pronoun *le* (singFEM) becomes *gli* when followed by another clitic, which results in homonymy and consequent breakdown of gender distinctions:

mostrale la stanza 'show her the room'

mostragli la stanza 'show him the room'

mostragliela 'show it to her/him'

Homonymy occurs also with other unstressed forms (*ci* and *vi* are also locative adverbs; *lo, la, gli, and le* articles) and with stressed pronouns in postverbal position (*te lo prendi con te* 'you take it/him with you').

Verb Inflection

The verb is the word class with the richest inflection both in number of forms and variability (many high-frequency verbs have irregular inflection). Tense, mood, person, and number are expressed affixally.

The most relevant features of the fusional type appear even within regular inflection. Grammatical meanings can be expressed as fused in a single morph (in the form *amo*, 'I love,' there is no other overt expression than the ending *-o* to code indicative mood, present tense, first-person, singular), as well as realized through a combination of several affixes, with complex exponence relations holding between morphs: in *canterébbero* '[if] they would sing' three morphemes (conditional, third person and plural) are signaled by the entire termination *-rébbero* as a whole. A linear segmentation is not possible because *-bb-* occurs only with forms that are both third person and conditional. The final *-ro* expresses plural and again third person, the stressed *-e-* occurs consistently

Table 3 Stressed personal pronouns

| Stressed Pronouns | 1 sing | 2 sing | 3 sing | 1 PL | 2 PL | 3 PL |
|-------------------|--------|--------|--|------|------|------------------------|
| Subject | io | tu | +hum: egli (MASC) ella (FEM) –hum: esso (MASC) essa (FEM) | noi | voi | essi (MASC) esse (FEM) |
| Nonsubject | me | te | +hum: lui (MASC) lei (FEM) –hum: esso (MASC) essa (FEM) | noi | voi | loro |

Table 4 Clitic personal pronouns

| Clitic Pronouns | 1 sing | 2 sing | 3 sing | 1 PL | 2 PL | 3 PL |
|-----------------|--------|--------|---------------------|------|------|--------------------|
| ACC | mi | ti | lo (MASC) la (FEM) | ci | vi | li (MASC) le (FEM) |
| DAT | mi | ti | gli (MASC) le (FEM) | ci | vi | - |

Table 5 Indicative present regular inflection of the three traditional verbal classes

| INDIC PR | Class-a-sing | PL | Class-e-sing | PL | Class-i-sing | PL |
|----------|--------------|-----------|--------------|----------|--------------|----------|
| 1 | cànt-o | cànt-iàmo | tèm-o | tem-iàmo | àpr-o | apr-iàmo |
| 2 | cànt-i | cànt-à-te | tèm-i | tem-é-te | àpr-i | apr-i-te |
| 3 | cànt-a | cànt-ano | tèm-e | tèm-ono | àpr-e | àpr-ono |

only in conditional forms, and the *-r-* occurs with futures and (again) conditionals (Matthews, 1970).

A suprasegmental modification (stress on the last syllable) suffices to distinguish between two regular verb forms like *tème* (INDIC PRES 3sing) and *temé* (INDIC preterit 3sing).

Table 5 shows instances of regular inflection of indicative present for the three traditional classes on the basis of the thematic vowel (accent marks have been added, even though not in use in normal orthography). Noteworthy, are the cumulative expression of more meanings by means of a single ending, and the omission of the thematic vowel with the exception of the second plural.

The thematic vowel may also undergo some allomorphisms: in class *-a-* it is replaced by /e/ in the indicative future (*amerò*) and the present conditional (*amerei*), in class *-i-* by /e/ in the present participle (*aprente*) and in the gerund (*aprendo*), in class *-e-* by /u/ in the past participle (*voluto*), and by /i/ in the derivation of deverbal nouns and adjectives (*temibile*, *spremitura*).

The indicative imperfect is the tense that least identifies with the fusional type, for it shows the best correspondence between morphs and morphemes in Italian conjugation (Table 6). But also in the imperfect the categories of person and number, as well as those of mood and tense, merge in a single morph. The person–number affixes largely overlap (even though not entirely) with those of the present, whereas the morph *-v-* has no other use in the conjugation.

Table 6 Indicative imperfect inflection of regular class *-a-* verbs

| INDIC Imperfect class -a- | sing | PL |
|---------------------------|------------|--------------|
| 1 | cant-à-v-o | cant-a-v-àmo |
| 2 | cant-à-v-i | cant-a-v-àte |
| 3 | cant-à-v-a | cant-à-v-ano |

Besides synthetic forms, there are also analytic forms resulting from the combination of an auxiliary and a past participle. The main auxiliary verbs are *essere* ‘to be’ and *avere* ‘to have’; both are highly irregular

INDIC PRES *sono, sei, è, siamo, siete, sono*
 INDIC preterit *fui, fosti, fu, fummo, foste, furono*
 INDIC IMPERF 1PERS *ero*, INDIC FUT 1PERS *sarò*;
 PP *stato*

The auxiliary expresses tense, mood, person, and number, whereas the past participle can either agree in gender and number or be employed in the citation form (singMASC), whereas the agreement between participle and subject is systematic with verbs whose auxiliary is *essere* (*i ragazzi sono partiti* ‘the boys have left’/ *la ragazza è partita* ‘the girl has left’), in verbs whose auxiliary is *avere* the agreement is with the object, and it occurs only within limited contexts (*hai comprato le pere? sì le ho comprate* ‘have you bought the pears? Yes, I have bought them’).

The indicative future and the present conditional are interesting from a typological point of view. They

have stemmed from an analytic origin to reach a fusional state.

Many verbs have an irregular conjugation. The majority belong to the *-e-* class, some others to the *-i-* class, only three to the *-a-* class. The traditional classification does not allow to catch the similarities between *-e-* and *-i-* classes. Dressler *et al.* (2003) classify Italian verbs in two macroclasses based on the productivity criterion, as well as on formal correlations that allow us to also take into account the subdivisions within each class.

It would be neither possible nor helpful to mention here all the instances in which the verbs depart from the regular conjugation model. Idiosyncrasies occur primarily in the present (indicative and subjunctive), the preterit, and the past participle.

Pirrelli and Battista (2000) show that in the irregular conjugations the modifications of the stems range along a continuum from minor phonological processes to clear suppletion instances.

Table 7 displays the phonetic transcription of indicative present conjugation of six irregular verbs that exhibit a number of stem alternations: palatalization before a front vowel (*nascere*), ablauting of root vowel (*udire*), diphthongization (*sedere*), *-isc-* insertion and palatalization (*finire*), ablauting, consonant labialization and lengthening (*dovere*), and stem suppletion (*andare*).

Two things are noteworthy here. The first is that phonological phenomena responsible for the stem modifications are synchronically inoperative. The second is that these different phenomena are distributed according to a recurrent pattern – which is visible also in the stress position within the regular conjugation: on the one hand 1,2,3sing 3PL on the other 1,2PL (Vincent, 1988).

The most common irregular modifications of preterit stems concern lengthening of final consonant (*venire*), replacing of final consonant with /s/ (*perdere*), and other more complex phenomena that include ablauting of root vowel and deletion/insertion of consonants (*fondere*). Even in the preterit a morphological pattern of stem alternation is detectable, which distinguishes 1,3sing 3PL vs. 2sing 1,2PL (see Table 8).

Even though the majority of alterations were originally phonologically motivated, their present distribution is morphological in nature and paradigmatically governed.

Affixes

Obligatoriness, higher systematicity in mutual relationships, and degree of productivity clearly distinguish inflectional from derivational affixes in fusional languages. Yet the number of productive Italian derivational affixes and the variety of meanings they express are fairly high (Grossmann and Rainer, 2004). Whereas derivation employs both suffixes and prefixes, inflection employs suffixes only. In productive processes the degree of phonological fusion between the stem and derivational affixes is rather low. A higher degree of phonological integration is present instead with interfixes. Some interfixes are semantically void (*congress-u-ale* ‘concerning congresses’), the majority express connotative value when combined with evaluative suffixes (*test-ol-ina* ‘nice small head’). Words containing interfixes are rather difficult to be segmented. This is because interfixes occupy an intermediate position between the root and the suffix, and generally their morphemic and syllabic boundary does not coincide.

Table 7 Examples of stem modifications in the indicative present inflection of some very frequent irregular verbs

| | <i>nascere</i> | <i>‘be born’</i> | <i>udire</i> | <i>‘hear’</i> | <i>sedere</i> | <i>‘sit’</i> | <i>finire</i> | <i>‘finish’</i> | <i>dovere</i> | <i>‘must’</i> | <i>andare</i> | <i>‘go’</i> |
|---|----------------|------------------|--------------|---------------|---------------|--------------|---------------|-----------------|---------------|---------------|---------------|-------------|
| | <i>sing</i> | <i>PL</i> | <i>sing</i> | <i>PL</i> | <i>sing</i> | <i>PL</i> | <i>sing</i> | <i>PL</i> | <i>sing</i> | <i>PL</i> | <i>sing</i> | <i>PL</i> |
| 1 | 'nasko | na'f:amo | 'ɔdo | u'djamo | 'sjedo | se'djamo | fi'nisko | fi'njamo | 'devo | do'b:jamo | 'vado | an'djamo |
| 2 | 'naʃ:i | na'f:ete | 'ɔdi | u'dite | 'sjedi | se'dete | fi'nij:i | fi'nite | 'devi | do'vete | 'vai | an'date |
| 3 | 'naʃ:e | 'naskono | 'ɔde | 'ɔdono | 'sjede | 'sjedono | fi'nij:e | fi'niskono | 'deve | 'devono | 'va | 'vanno |

Table 8 Examples of stem modifications in the indicative preterit inflection of some very frequent irregular verbs

| | <i>venire</i> ‘come’ | | <i>perdere</i> ‘lose’ | | <i>fondere</i> ‘melt’ | |
|---|----------------------|-----------|-----------------------|-----------|-----------------------|-----------|
| | <i>sing</i> | <i>PL</i> | <i>sing</i> | <i>PL</i> | <i>sing</i> | <i>PL</i> |
| 1 | vénni | venimmo | pèrsi | perdémmo | fùsi | fondémmo |
| 2 | venisti | veniste | perdesti | perdeste | fondèsti | fondèste |
| 3 | vénne | vénnero | pèrse | pèrsero | fùse | fùsero |

The clear-cut difference between affixes and roots characterizes Italian as a fusional language and distinguishes it from agglutinating languages, which grant affixes greater autonomy (Plungian, 2001). In Italian the coordination between bound elements referring to a same root is marginal (*micro- e macro-economia* 'micro- and macroeconomy'); moreover, there is no use of a single affix referring to a coordinated group (cf. Spanish *afirmativa o negativamente* 'affirmatively or negatively'). Another fusional trait of Italian derivation is the signaling through the change of inflectional class of derivatives formed by means of conversion, which belong to the same word class as the base (*banana* FEM, PL *-e* 'banana' → *banano* MASC, PL *-i* 'banana tree').

Word Order Mobility

The Italian basic word order (SVO) is more flexible than what we encounter in less-fusional languages. Agreement (in gender and number) provides cohesion among words within phrases, and a certain degree of mobility freedom for the phrases themselves. Instead, single words are less free to move compared to words in more inflecting languages, such as Latin, for instance, (Simone, 1993). Whereas in Latin, the use of cases allows to signal on each word its relational syntactic functions, thus rendering it relatively autonomous within the phrase, in Italian the analytic expression (through prepositions) of syntactic relations demands the proximity and the reciprocal ordering of words within the phrases. Thus, order variation occurs primarily at the level of the reciprocal ordering of phrases (many instances of topicalization, cleft sentences, postverbal subject position). Adjectives stand out from other word classes for a higher degree of movement freedom within the phrases. Even though the unmarked position of the qualifying adjective is postnominal (according to the basic order SVO), this may vary, which at times effects a change

in meaning (cf. *famiglie numerose* 'large families' and *numerose famiglie* 'several families').

Bibliography

- D'Achille P & Thornton A M (2003). 'La flessione del nome dall'italiano antico all'italiano moderno.' In Maraschio N & Poggi Salani T (eds.) *Italia linguistica anno mille Italia linguistica anno duemila*. Rome: Bulzoni.
- Dressler W U (1985). 'Typological aspects of natural morphology.' *Acta Linguistica Academiae Scientiarum Hungaricae* 35, 51–70.
- Dressler W U, Kilani-Schoch M, Spina R & Thornton A M (2003). 'Le classi di coniugazione in italiano e in francese.' In Giacomo-Marcellesi M & Rocchetti A (eds.) *Il verbo italiano*. Rome: Bulzoni.
- Greenberg J H (1954). 'A quantitative approach to the morphological typology of language.' In Spencer R F (ed.) *Method and perspective in anthropology*. Minneapolis: University of Minnesota Press.
- Grossmann M & Rainer F (eds.) (2004). *La formazione delle parole in italiano*. Tübingen: Niemeyer.
- Matthews P H (1970). 'Recent developments in morphology.' In Lyons J (ed.) *New horizons in linguistics*. Harmondsworth: Pelican.
- Pirrelli V & Battista M (2000). 'The paradigmatic dimension of stem allomorphy in Italian verb inflection.' *Italian Journal of Linguistics* 12, 307–379.
- Plungian V A (2001). 'Agglutination and flection.' In Haspelmath M, König E, Oesterreicher W & Raible W (eds.) *Language typology and language universals*. Berlin, New York: De Gruyter.
- Skalička V (1966). 'Ein "typologisches Konstrukt."' In Skalička V (ed.) (1979). *Typologische Studien*. Braunschweig: Vieweg.
- Simone R (1993). 'Stabilità e instabilità nei caratteri originali dell'italiano.' In Sobrero A (ed.) *Introduzione all'italiano contemporaneo*. Rome-Bari: Laterza.
- Thornton A M (2001). 'Some reflections on gender and inflectional class assignment in Italian.' In Schaner-Wolles C et al. (eds.) *Naturally!* Turin: Rosenberg & Sellier.
- Vincent N (1988). 'Italian.' In Harris M & Vincent N (eds.) *The Romance languages*. London: Routledge.

Italic Languages

D Ridgeway, University of Edinburgh, Edinburgh, UK

© 1994 Elsevier Ltd. All rights reserved.

The adjective 'Italic' is conventionally applied to a group of related Indo-European languages attested epigraphically in the Italian peninsula during the later first millennium BC. Their written forms virtually

disappear from the archaeological record by the end of the first century BC following the rise of Latin, which, together with the sparse remains of Faliscan (in the area bordering on non-Indo-European Etruria), constitutes one of two Italic sub-groups; the other (for some authorities the only true Italic category) is represented along the spine of Italy by Oscan, Umbrian, and associated minor dialects.

Elsewhere, Messapic is a non-Italic branch of Indo-European in Apulia; further up the east coast, the same is increasingly thought to be true of Picene ('East Italic') and Venetic.

That the Italic languages are the result of intrusion is not in doubt. The traditional view of Indo-European invaders sweeping across the Alps has been superseded by a complex model involving linguistic innovation in the south and long periods of cohabitation and fusion between indigenous prehistoric communities and successive groups of immigrants. Further internal population movements were well advanced when speakers of the Italic and North Italian Indo-European languages developed their own versions of the Etruscan alphabet.

Around 300 inscriptions in Oscan, many of them very short, occur from ca. 400 BC principally in Campania (where graffiti at Pompeii show that it was still occasionally written as late as 79 AD), and to a lesser extent in ancient Lucania and Bruttium (modern Basilicata and Calabria). Roman antiquarians defined Oscan speakers as *Sabelli* (synonymous with 'enemies of Rome'); historians prefer 'Sabellian' for the Samnite and other speakers of Oscan proper, and 'Sabellic' for the speakers of the Oscan-type dialects occasionally represented between the third and first centuries BC in the mountains of central Italy (Paelignian, Vestinian, Marrucinian, and Marsian). The longest Oscan text is inscribed on the bronze Tabula Bantina, found in 1793 on the boundary between Apulia and Lucania; its six paragraphs, part of a much longer document, retail municipal regulations of the early first century BC and are written in the Latin alphabet. Most of the other texts are inscribed in the Oscan alphabet; a few use Greek letters.

Apart from two dozen short inscriptions from the fourth century BC onwards, the Umbrian language is known entirely from the texts inscribed in the Umbrian and Latin alphabets on the Tabulae Iguvinae: seven bronze tablets, containing over 4000 words, discovered in 1444 at Gubbio. Written at intervals between the late third and late second centuries BC, they record the proceedings of a priestly college, the *frater atiiēiur* ('Atiedian Brethren'), and constitute

the largest pre-Christian liturgical corpus in Europe. Two dialects, Aequian and Volscian, show clear affinities with Umbrian; they are represented in parts of Latium known to have been settled by their speakers in the early fifth century BC.

There is nothing literary about the extant remains of any of the Italic languages except Latin. The substance of those noted above is of interest principally to historians of Italic religious and political institutions during the centuries of Roman hegemony. Although the morphology and syntax of Oscan and Umbrian inevitably have much in common with their Latin counterparts, significant differences survive; they include a third person singular passive subjunctive in *-r* and the extensive use of the locative case: Oscan *sakrafir; eísaí víaí* ('let there be sacrifice of'; 'on that road'); Umbrian *ferar; destre onse* ('let it be carried'; 'on the right shoulder'). References by Roman writers to *meddix tuticus* as 'chief magistrate of the people' are matched by the occurrence of *meddíss túvtiks* and numerous variants in Sabellian and Sabellic contexts. Writing in the seventeenth century, the Scottish classical scholar Thomas Dempster wrongly took the term to be Etruscan, and cited it in connection with a wholly fanciful ancient origin for the family of his Medici patron in Florence.

Bibliography

- Buck C D (1928). *Grammar of Oscan and Umbrian*, 2nd edn.). Boston, MA: Ginn; reprinted 1979, Georg Olms, Hildesheim and New York.
- Marinetti A (1984). *Lingue e dialetti dell'Italia antica: aggiornamenti ed indici = Popoli e civiltà dell'Italia antica* (vol. 6-bis). Rome: Biblioteca di Storia Patria.
- Pallottino M (1991). *A History of Earliest Italy* (transl. Ryle M, Soper K). Ann Arbor, MI and Routledge, London: University of Michigan Press.
- Penney J H W (1988). The languages of Italy. In *Cambridge Ancient History*, 2nd edn., vol. 4. Cambridge: Cambridge University Press. 72038.
- Prosdocimi A L (ed.) (1978). *Lingue e dialetti dell'Italia antica = Popoli e civiltà dell'Italia antica* (vol. 6). Rome: Biblioteca di Storia Patria.

This page intentionally left blank

J

Japanese

M Shibatani, Rice University, Houston, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

Japanese is spoken by virtually the entire population of Japan – some 128 million people in 2004. In terms of native speakers, the number easily surpasses the number of speakers of the major European languages (German and French), ranking sixth among the languages of the world, after Chinese, English, Russian, Hindi, and Spanish. Yet, among the major languages of the world, Japanese occupies a unique position in a number of respects. Unlike the languages spoken on the European, American, and Asian continents, the Japanese language, being spoken in an island nation, is physically isolated from other languages. Also, unlike major European languages such as English and Spanish, Japanese is primarily spoken within the confines of its national boundaries, with no other country using it as either a first or second language. Moreover, Japanese is the only major world language for which the genetic affiliation to other languages and language families has not been conclusively proved.

As in the case of other language isolates, including Basque and Burushaski (and the geographical neighbors of Japanese, Ainu and Korean) (*see Ainu and Korean*), the genealogy of Japanese has been a perennial problem that has attracted the attention of both specialists and laymen alike. Hypotheses have been presented assigning Japanese to virtually all major language families. Although attempts to relate Japanese to the Altaic family have been most systematic and perhaps most persuasive (*see Altaic Languages*) two views have attracted increasing attention in recent years: (1) that Japanese consists of an Austronesian substratum and an Altaic superstratum, and (2) that Japanese is an Austronesian–Altaic hybrid or mixed language, in which not only simple lexical mixtures but also morphological hybrids, e.g., Austronesian verb roots with Altaic inflectional endings, are recognized (*see also Austronesian Languages*).

Among individual languages, Ryukyuan, Ainu, and Korean are the strongest candidates proposed as

possible sister languages; in fact, the Japanese–Ryukyuan connection has been firmly established, and Ryukyuan, spoken in Okinawa, is now considered to be a dialect of Japanese (*see Ryukyuan*). A Japanese–Ainu relationship has been hypothesized, but evidence is scanty. There have been attempts to make more systematic comparisons between Japanese and Korean, but even here reliable sets of cognates are extremely small in number.

Being a mountainous country with numerous islands, Japan is an ideal setting for fostering language diversification. Indeed, Japanese is extremely rich in dialectal variations, and different dialects are often mutually unintelligible. However, Japan is linguistically completely unified by a uniform writing system and by the spread of the standardized speech, based on the Tokyo dialect, making oral communication possible among speakers of different dialects. The long literary history of Japanese, which dates back to the 8th century, finds its root in the borrowing of Chinese characters as a means of transcribing Japanese. Simplification of Chinese characters gave rise to two kinds of syllabary, or kana, i.e., *hiragana* and *katakana*. In addition to Chinese characters and the two types of syllabary, the Japanese writing system includes the Roman alphabet, which was introduced in the late 16th century by Portuguese and Spanish missionaries. All four types of writing systems have been retained today, and it is not unusual to see contemporary Japanese sentences written with a mixture of all of them.

The historical contacts with foreign cultures have left strong marks in the Japanese lexicon, which is characterized by a high percentage of loanwords. Roughly 60% of the Japanese vocabulary consists of loanwords of Chinese origin, a figure comparable to the proportion of Latinate words in the English vocabulary. Among the non-Chinese loanwords, or roughly 10% of Japanese vocabulary, English loans stand out, often replacing older loans from Portuguese, Spanish, and Dutch as well as even some Chinese loans.

Another salient feature of the Japanese lexicon is the presence of a large number of sound symbolic or

mimetic words, which depict not only the sounds of natural objects and animals but also the manners of action and even states of mind. Thus, a dog barks *wan-wan* and it pours rain *zaa-zaa*; an old man walks *yobo-yobo* ‘wobbly’ and an old lady chatters *peya-kutya*; and your head aches *zuki-zuki* ‘throbbingly,’ your stomach hurts *tiku-tiku* ‘stingingly,’ and your nerves are irritated *ira-ira*.

Segmental phonology of Japanese – at least the speech of Tokyo and surrounding areas – is rather simple, with five vowel phonemes /a e i o u/ and 16 consonantal phonemes /p t k b d g s h z r m n w j N Q/. A noteworthy feature of Japanese segmental phonology is the distinction between a syllable and a mora. A mora is a unit that can be represented by one letter of kana (a Japanese pseudo-character used in syllabic writing). A word such as *simbun* ‘newspaper’ consists of two syllables, but a Japanese speaker further subdivides the word into four units, *si*, *m*, *bu*, and *n*, which correspond to the four letters of kana used in the written form. The consonantal archiphonemes /N/ and /Q/ correspond to moraic consonants seen in words such as *simbun* ‘newspaper,’ *simpai* ‘worry,’ *hakkiri* ‘clearly,’ and *kossori* ‘stealthily,’ wherein the first segments of the consonant clusters, *m*, *k*, and *s*, constitute moras. Since these moraic consonants are homorganic to the following consonants, there is no contrast among them other than in terms of the nasality feature. And since the only consonant that ends a word is the moraic nasal *n*, the words *simbun*, *simpai*, *hakkiri*, and *kossori* are phonemicized, respectively, as /siNbuN/, /siNpai/, /haQkiri/, and /koQsori/. The units of syllable and mora each play important roles in the accentual system.

Japanese is arguably a tone language, wherein the pitch pattern can be high followed by low (HL) or low followed by high (LH). Pitch height alone distinguishes minimal pairs such as *hasi* (HL) ‘chopstick’ versus *hasi* (LH) ‘bridge’ and *ame* (HL) ‘rain’ versus *ame* (LH) ‘candy’; however, the Japanese accentual system is characteristically different from archetypal tone languages of the Chinese type, in which it is necessary to specify the tone for each syllable. In the case of (Tokyo) Japanese, once the place of pitch drop for a given word or a minor phonological phrase is specified, the pitch shape can be fully predicted. Thus, *hasi* ‘chopstick’ and *hasi* ‘bridge’ can be specified as /ha’si/ and /hasi’/, respectively, so that in the former, high pitch drops after *ha*, whereas in the latter, all moras are high except for the initial one. A pitch drop in the latter form is observed only when it is followed by another element, such as the nominative particle *ga*, as in the minor phrase /hasi’ga/ (LHL) ‘bridge NOM.’ The word *hasi* (LH) ‘edge’ is representable as

/hasi/ without any accent; indeed, there is no pitch drop here, even in a phrase such as /hasi ga/ (LHH) ‘edge NOM.’ Pitch changes occur at mora boundaries, and the accent marker indicating the location of pitch drop is assigned to the unit of syllable. For example, the three-mora, two-syllable word *ganko* ‘stubborn’ contains the accent marker in the initial syllable, /ga’Nko/, surfacing with the pitch shape of HLL. There is no word in which the second mora of the first syllable carries the accent, such as /gaN’ko/, which would presumably be pronounced as LHL or HHL if the syllable were the unit of tone inflection.

Japanese is an agglutinative language with primarily suffixing morphology. Both verbs and adjectives inflect for tense, but they are distinguished by different tense suffixes: *mi-ru* ‘see-PRES’ and *mi-ta* ‘see-PAST’ versus *utukusi-i* ‘beautiful-PRES’ and *utukusi-kat-ta* ‘beautiful-EXPL-PAST’ (EXPL = expletive). In addition to the inflecting adjectives, adjectival nominals are similar in meaning to adjectives but, like nominal predicates, call for tense-carrying copula (COP) in their predicative function: e.g., *kirei-da* ‘pretty-COP-PRES’ and *kirei-da-tta* ‘pretty-COP-PAST.’ These tense suffixes combine with various auxiliary-type suffixes, often resulting in a fairly long verbal complex: *ika-se-rare-ta-gara-na-i* (go-CAUS-PASS-DESI-SHOW-NEG-PRES) ‘do not show signs of wanting to be made to go.’

Japanese syntax is consistently head-final. The basic word order is subject-object-verb: *Taroo ga hon o yonda* (Taro NOM book ACC read-PAST) ‘Taro read a book.’ Postpositional particles are used instead of prepositions, as in the example, wherein the nominative and accusative (ACC) particles *ga* and *o*, respectively, mark the subject and the object. Modifiers precede the heads that they modify: *takai hon* (expensive book) ‘expensive book’ [*Taroo ga katta*] *hon* ([Taro NOM bought] book) ‘the book Taro bought,’ *Taroo no ie* (Taro GEN house) ‘Taro’s house,’ *sono hon* (that book) ‘that book,’ *san-satu no hon* (three-CLASS GEN book) ‘three books,’ *hayaku hasiru* (quickly run) ‘run quickly,’ *tabe-tai* (eat-want) ‘want to eat,’ *Taroo yori kasikoi* (Taro than smart) ‘smarter than Taro.’ Subordinating conjunctions occur after subordinate clauses, which in turn come before main clauses: [*Taroo ga kita*]-*node minna ga kaetta* ([Taro NOM came]-because everyone NOM went home) ‘Because Taro came, everyone went home.’

One of the most important aspects of Japanese grammar has to do with the topic construction. The topic particle *wa* attaches to various nominals and adverbials, yielding topic sentences that contrast with nontopic sentences in the following manner:

Nontopic sentence (1):

- (1) Taroo ga Ziroo no hon o yonde-iru
Taro NOM Jiro GEN book ACC read-be
'Taro is reading Jiro's book'

Topic sentences (2) and (3):

- (2) Taroo wa Ziroo no hon o yonde-iru
Taro TOP Jiro GEN book ACC read-be
'Taro is such that he is reading Jiro's book'
- (3) Ziroo no hon wa Taroo ga yonde-iru
Ziroo GEN book TOP Taro NOM read-be
'Jiro's book is such that Taro is reading it'

The basic difference between topic sentences and nontopic sentences is that the former are statements about certain things, represented by topic nominals, and the latter are statements describing the occurrences of events. For example, nontopic sentence (1) describes the event of Taro reading Jiro's book. On the other hand, topic sentences (2) and (3), respectively, describe something about 'Taro' and 'Jiro's book.' Thus, sentence (1) answers a question such as 'What is happening?,' whereas sentences (2) and (3), respectively, would be used in answering questions such as 'What is Taro doing?' and 'Where is Jiro's book?' In Japanese, it is the *wa*-marked topic nominal that accurately represents the traditional Western definition of subject as something that is being talked about. The *ga*-marked subject nominal, on the other hand, is more consonant with the other definition of subject, namely, that it expresses an actor or agent. In other words, in Japanese, the two notional definitions of subject are distributed over two distinct syntactic relations, whereas in English and other European languages, these largely converge on the single subject nominal.

Japanese has no agreement marker, but it freely omits pronouns. Thus, the following type of exchange is not uncommon:

- (4) *Oo kita ka* (Oh came Q) 'Oh (you) came?'
- (5) *Un kita* 'Yeah (I) came'

Of course, the omissions of pronouns are permitted only when they are recoverable from the context; one type of clue for the recovery is found in the honorific endings. Humbling forms together with the polite (POL) ending such as *Mairimasu* (go.HUMBLE-POL) 'go' and *O-tazune simasu* (HON-visit do-POL) 'visit' indicate (a) that the subject is either a speaker or someone close to the speaker (by use of the humbling forms) and (b) that the addressee is someone worthy of respect (by use of the polite ending). On the other

hand, honorific forms with the plain ending, such as *Oide-ni naru* (go.HON-ADV become.PLAIN) 'go' and *Otazune-ni naru* (HON-visit-ADV become.PLAIN) 'visit' indicate (a) that the subject is other than the speaker and is someone worthy of respect (honorific forms) and (b) that the addressee is someone close to the speaker, to whom the speaker is not obliged to show respect (plain ending). Thus, both the addressee axis (for polite endings) and the referent axis (for honorific and humbling endings) control honorific phenomena independently, though they often converge, as when the addressee is also the referent of the subject nominal.

Another clue that helps identify the nature of the speaker is sentence final particles, some of which are different for male and female speakers. The final particle (PART) *wa* is a typical female form, whereas *zo* occurs in rough male speech. Since these discourse particles occur in intimate speech, an expression such as *Mairimasu wa* (go.HUMBLE-POL PART) '(I will) go' indicates that the subject is the speaker (humble form), that the addressee is someone worthy of respect (polite ending), and that the speaker is a woman who is on intimate terms with the addressee (final particle).

Thus, Japanese, though it lacks agreement markers, has a number of grammatical features that not only indicate the nature of the subject but also index the social relationships between the speaker and the addressee and between the speaker and the nominal referent, as well as the gender of the speaker. These features, on the other hand, require the speaker of Japanese to predetermine the social relationships between the speaker and the addressee and the nominal referent, so that appropriate combinations of honorifics and discourse particles can be chosen. Japanese, in other words, is a highly context-sensitive language in which individual expressions encode various factors that make up conversational contexts in which they are embedded.

Bibliography

- Kuno S (1973). *The structure of the Japanese language*. Cambridge, MA: MIT Press.
- Martin S E (1975). *A reference grammar of Japanese*. New Haven, CT: Yale University Press.
- McCawley J D (1968). *The phonological component of the grammar of Japanese*. The Hague: Mouton.
- Miller R A (1971). *Japanese and the other Altaic languages*. Chicago, IL: University of Chicago Press.
- Shibatani M (1990). *The languages of Japan*. Cambridge: Cambridge University Press.

Javanese

B Nothofer, Universität Frankfurt, Frankfurt, Germany

© 2006 Elsevier Ltd. All rights reserved.

Javanese is most-spoken regional language of Indonesia and most-spoken language of the Austronesian language family with about 75 million speakers. It is spoken along the northwest coast of Java (Banten, Krawang, Cirebon) and in the central and eastern areas of this island (68 million speakers). Outside of Java, it is used in the Indonesian transmigration areas of Sumatra, Kalimantan, and Sulawesi (altogether 8 million speakers) as well as in Suriname (60 000 speakers) and in New Caledonia (6700 speakers). Of the three dialects usually distinguished (western, central, and eastern) (Ras, 1994), the western one was divided into seven subdialects (Nothofer, 1980). The central and eastern dialect variants have not yet been studied in much detail.

Standard Javanese is the language as it is spoken in the area of Surakarta and Yogyakarta. Javanese has speech levels that are based on the principle to whom and about whom one talks. The level chosen depends on factors such as age, status, and respect. The degree of politeness is expressed by lexical or affixal choices. The speech levels are a Javanese innovation. Clynes (1992) argues that this system was well established by the 15th century. Sundanese (Sunda), Madurese (Madura), Balinese (Bali), and Sasak (but not Malay) have borrowed these speech levels. The position of Javanese in the Western Malayo-Polynesian subfamily has been a matter of dispute. While Dyen (1965) and Nothofer (1975) grouped Javanese with Malay, Sundanese, and Madurese, Nothofer (1985) suggests that the latter three – although not necessarily constituting a subgroup – are more closely related to each other than they are to Javanese. Javanese is one of the few Austronesian languages whose history can be traced because of the existence of older texts. The oldest records date back to the 8th century. Zoetmulder (1974) deals with Old Javanese literature and Zoetmulder (1982) is an Old Javanese-English dictionary.

Javanese Phonology

Consonants

The Javanese consonant system resembles that of other languages of western Indonesia. There are features that are common to Javanese and Madurese only. Both languages share a phonemic distinction between dental and retroflex stops. The ‘voiced’ consonants of Javanese are pronounced like voiceless stops with breathy voice of the following vowel (Fagan, 1988; Arps *et al.*, 2000). The consonants are shown in Table 1.

Vowels

Nothofer (1980) suggests a system of six vowel phonemes, shown in Table 2.

Javanese has the following allophonic rules (Clynes, 1995; Nothofer, 1980): /i/, /u/ are realized as [ɪ], [ʊ] in closed syllables and as [i], [u] elsewhere. /ə/ is always realized as [ə]. The phonemes /e/, /o/ are realized as [ɛ] and [ɔ] in closed syllables, in open syllables where the vowel in a following open syllable is high, and in open syllables where the vowel in a following syllable is identical or /ə/. In all other positions, these phonemes are realized as [e] and [o]. The phoneme /a/ appears as [ɔ] word-finally and in penultimate open syllables where a following open syllable has /a/. Otherwise, it appears as [a]. Allophonic variation also depends on the initial phoneme of suffixes: the addition of a consonant-initial suffix results in the treatment of the stem-final vowel as if it appeared in a closed syllable and the addition of a vowel-initial suffix will cause a high vowel in the stem-final closed syllable to behave as if it appeared in an open syllable.

Morphology

Verbal affixes include the following: N- indicates an ‘active’ and *di-* a ‘passive’ transitive verb, *kə-* marks ‘accidental passive.’ *-ke* forms transitive verbs whose patient is the causee or benefactee, while *-i* forms transitive verbs in which the undergoer is the

Table 1 Javanese consonants

| Consonant type | Labial | Dental | Retroflex | Palatal | Velar | Glottal |
|-----------------|----------|----------|-------------|----------|----------|----------|
| Voiceless stops | <i>p</i> | <i>t</i> | <i>ʈ</i> | <i>c</i> | <i>k</i> | <i>ʔ</i> |
| ‘Voiced’ stops | <i>b</i> | <i>d</i> | <i>ɖ</i> | <i>j</i> | <i>g</i> | |
| Nasal | <i>m</i> | | <i>n</i> | <i>ɲ</i> | <i>ŋ</i> | |
| Fricative | | | <i>s</i> | | | <i>h</i> |
| Approximant | <i>w</i> | | <i>r, l</i> | <i>y</i> | | |

Table 2 Javanese vowel phonemes

| Vowel type | Front | Central | Back |
|------------|----------|----------|----------|
| High | <i>i</i> | | <i>u</i> |
| Mid | <i>e</i> | <i>ə</i> | <i>o</i> |
| Low | | <i>a</i> | |

location or goal of the action. Intransitive verbs with *-an* indicate ‘nonchalance, to be in the state of, to busy o.s. with.’ The suffix *-a* with the sense ‘in case, assuming, although’ occurs with nonverbs (e.g., pronouns, adjectives) and verbs. Arps *et al.* (2000) call it the ‘irreality-suffix.’ It is distinguished from the verbal suffix *-a*, which is an imperative marker. The suffix *-ən* is added to nouns denoting physiological conditions and forms verbs meaning ‘undergo the physiological process of (the noun).’ The nominal affixes include *-an*, which derives nouns that stand as objects of an action indicated by the verb; *kə-* *-an*, used to nominalize qualities; nouns formed with *pəN-* refer to a person carrying out the action of the verb or an instrument with which the action is performed; and *pə-an* nouns refer to a process or result of an action indicated by the verb or to the location where an action of the corresponding verb occurs.

The meaning of total reduplication of nouns is ‘diversity, completeness,’ while that of verbs has the sense of ‘durative, intensive, iterative.’ Verbal reduplication can also involve vowel variation in the first member of the doubled form.

Writing System

After the World War II, the publication of texts in Javanese script (hanacaraka or caraka) came to an end. Actually, the Latin script began to replace the Javanese alphabet at the beginning of the 20th century. The traditional script originates from a Pallava script of southern India.

Bibliography

Arps B *et al.* (2000). *Semaian 20: Hedendaags Javaans. Een leerboek*. Leiden: Opleiding Talen en Culturen van Zuidoost-Azië en Oceanië, Universiteit Leiden.

Clynes A (1992). ‘Old Javanese influence on Balinese: speech styles.’ In Dutton T & Tryon D T (eds.) *Language contact and change in the Austronesian world*. Berlin, New York: Mouton de Gruyter.

Clynes A (1995). ‘Javanese.’ In Tryon D T (ed.) *An introduction to Austronesian studies 1: Fascicle 1: Comparative Austronesian dictionary*. Berlin, New York: Mouton de Gruyter. 469–484.

Dyen I (1965). *A lexicostatistical classification of the Austronesian languages*. IUPAL memoir 19; supplement to IJAL 25. Baltimore: Waverly Press.

Errington J J (1998). *Studies in the Social and Cultural Foundations of Language 19: Shifting languages: Interaction and identity in Javanese Indonesia*. Cambridge: Cambridge University Press.

Fagan J L (1988). ‘Javanese intervocalic stop phonemes: the light/heavy distinction.’ In McGinn R (ed.) *Studies in Austronesian linguistics*. Athens, Ohio: Center for Southeast Asia Studies, Ohio University.

Keeler W (1984). *International Studies, Southeast Asian Series 69: Javanese: A cultural approach*. Athens: Ohio University Center for International Studies.

Molen W van der (1993). *Semaian 8: Javaans schrift*. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit Leiden.

Nothofer B (1975). *Verhandelungen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 73: The reconstruction of Proto-Malayo-Javanic*. The Hague: Nijhoff.

Nothofer B (1980). *Dialektgeographische Untersuchungen in West-Java und im westlichen Zentral-Java*. Wiesbaden: Harrassowitz.

Nothofer B (1981). *Dialektatlas von Zentral-Java*. Wiesbaden: Harrassowitz.

Nothofer B (1985). ‘The subgrouping of the languages in the Javo-Sumatra Hesion: a reconsideration.’ *Bijdragen tot de Taal-, Land- en Volkenkunde* 141(2/3), 288–302.

Pigeaud Th (n. d.). *Nederlands-Javaans handwoordenboek*. Groningen: Wolters-Noordhoff.

Ras J J (1994). *Inleiding tot het modern Javaans*. Leiden: KITLV Uitgeverij.

Sudaryanto & Pranowo (eds.) (2001). *Kamus Pepak Basa Jawa*. Yogyakarta: Badan Pekerja Kongres Bahasa Jawa.

Sudaryanto *et al.* (1991). *Tata bahasa baku bahasa-Jawa*. Yogyakarta: Duta Wacana University Press.

Zoetmulder P J (1974). *Kalangwan: a survey of Old Javanese literature*. The Hague: Nijhoff.

Zoetmulder P J (1982). *Old Javanese dictionary* (2 vols). The Hague: Nijhoff.

Jèrriais

M C Jones, University of Cambridge, Cambridge, UK

© 2006 Elsevier Ltd. All rights reserved.

Background

Jèrriais is the dialect spoken on Jersey, the largest of the Channel Islands. It is related to the Norman dialects of northern France.

According to the 2001 Census of Jersey, there then remained only 2874 speakers of Jèrriais, (3.2% of the total resident population). Some two-thirds of these were over age 60 and only 113 speakers declared Jèrriais to be their usual everyday language.

Phonological Structure

Although the phonological structure of Jèrriais is similar to that of standard French, there are also significant differences. Marked regional variation is still very much in evidence in modern Jèrriais, with the sub-varieties usually categorized in two main groups: East and West Jèrriais (see Figure 1). Even more localized variation is readily observable, although with the decline in speaker-numbers, these so-called linguistic pockets are fast disappearing. This internal variation has never been based on any administrative or other territorial boundaries within Jersey, but many of the Islanders feel it to be intrinsically linked with parish boundaries (see Figure 2), and the practice of using the name of parishes to refer to the sub-dialects of Jèrriais is well established, although not strictly correct. Unless stated, the following comments relate to most varieties of Jèrriais.

Vowels:

Oral: /i e ε y œ a u o ə/; the French vowels /ø/ and /ɔ/ are lacking

Nasal: /ẽ ẽ̃ õ ã/;

Unlike in standard French, vowel length is phonemic and all vowels, except /ə/, can be either short or long. Long /a/ is usually realized phonetically as [a]. Jèrriais therefore has 27 vowel phonemes (17 oral, 10 nasal) compared to the 16 of standard French.

Consonants:

Stop: /p b t d k g/

Fricative: /f v s z ð ʃ ʒ h/

Affricate: /tʃ dʒ/

Nasal: /m n ɲ/

Lateral: /l/

Trilled: /r/; this corresponds to the uvular *r* of standard French

/ð/ (written *th*) occurs mainly as a result of the assimilation of intervocalic /r/ in western Jersey and some northern parts of Trinité and St. Martin e.g., *dithe* 'to say' (Fr. *dire*). In St. Ouen, /ð/ also occurs as a development of intervocalic /z/, e.g., *maïson* /meðõ/ 'house' (/me(j)zõ/ elsewhere).

In standard French, the affricates /tʃ/ and /dʒ/ only occur in borrowings, e.g., *match*, *gin*. However, in Jèrriais they can occur as the result of the secondary palatalization of /k/ and /g/, e.g., *tchoeu* /tʃœ/ 'heart', *dgèrre* /dʒɛ:r/ 'war' (Fr. *coeur*, *guerre*) and of /t/ and /d/, e.g., *mêtchi* /metʃi/ 'profession', *dgix* /dʒi/ 'ten' (Fr. *métier*, *dix*).

Like standard French, Jèrriais preserves no trace of Latin /h/. /h/ was introduced into the French phonemic system in words borrowed from the language of the Germanic invaders who dominated northern Gaul in the 5th to the 8th century. The sound disappeared

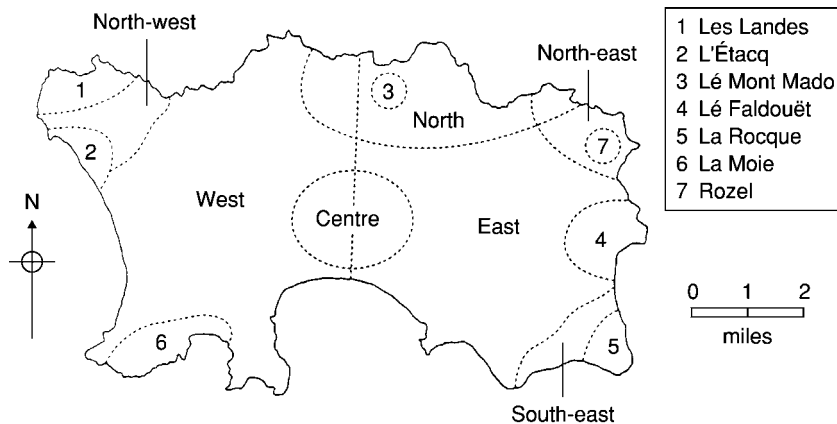


Figure 1 Jersey's 'linguistic pockets.' Reproduced from Jones M C (2001). *Jersey Norman French: a linguistic study of an obsolescent dialect*. Oxford: Blackwell.

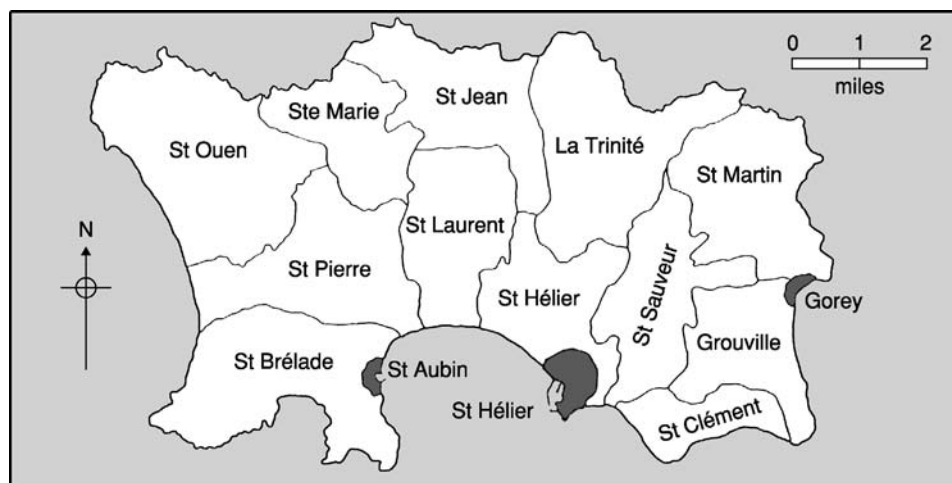


Figure 2 The twelve parishes of Jersey. Reproduced from Jones M C (2001). *Jersey Norman French: a linguistic study of an obsolescent dialect*. Oxford: Blackwell.

from standard French in the early modern period (16th–18th century) but remains in Jèrriais, e.g., *housse* /hus/ ‘holly’ (Fr. *houx*).

In many parts of Jersey, when /l/ is the second element of a consonant cluster, it frequently undergoes delateralization to /j/, e.g., *clios* /kjo/ ‘field’ (Fr. *clos*). In St. Ouen, however, the more conservative form /kʎo/ is still to be heard amongst older speakers. St. Ouen also retains the /k/ pronunciation word-finally in words such as /fiʔ/ *fil’ye* (‘daughter’) whereas, elsewhere in Jersey, /k/ has become depalatalized to /l/.

The velar nasal /ŋ/ occurs in standard French only in borrowings, e.g., *le shopping* /ʃɔpiŋ/. Despite its restricted distribution, it is generally given phonemic status. The sound also occurs in Jèrriais in English borrowings such as *blanket*, *dinghy* /blæŋket/, /dɪŋi/. However in Jèrriais /ŋ/ is not considered to be phonemic.

Glides The three glides of standard French (/j ɥ w/) also occur in Jèrriais but there is a tendency for /ɥ/ to be replaced by /w/.

Vocabulary

Most of the vocabulary of Jèrriais is shared with standard French, but regional variation is also apparent in the lexis of Jèrriais, e.g., *pêtre* ‘spider’ (WJ), *ithangnie* (EJ). The dominant influence of English has had far-reaching linguistic consequences, most noticeably in the lexis, where borrowings abound in many everyday domains. Some of these are well established, e.g., *bouchet* ‘bucket’, *tiç’ye* (‘tea’)kettle’ while others are more recent, e.g., *software*. The semantic adaptation of Jèrriais words on the basis of their English equivalents is also found,

e.g., *J’n’sais pon comment chenna travaille* ‘I don’t know how that works’, as are calques of English phrasal verbs and other expressions, e.g., *i’tchit bas* ‘he fell down’; *j’chèrчис pour* ‘I looked for it’.

Morphosyntax

The more striking morphosyntactic differences between Jèrriais and standard French include:

- i. the Old French distinction of number in masculine nouns and adjectives between *-el* (s.), *-eaus* (pl.), e.g., *chastel* ‘castle’, pl. *chasteaus*; *novel* ‘new’, pl. *noveaus* ‘new’, has been lost in standard French as a result of the creation of a new analogical singular based on the plural, e.g., *château* ~ *châteaux* (both pronounced [ʃato]), *nouveau* ~ *nouveaux*, though the original singular remains before vowels, e.g., *le nouvel an* ‘New Year’; in Jèrriais, the distinction is maintained, *châté* ~ *châtchieaux*, *nowé* ~ *nou-vieaux*;
- ii. the first person plural personal pronoun subject *nous* is replaced by *jélj*, e.g., *j’pâlons* (Fr. *nous parlons*) ‘we speak’;
- iii. there is no specific feminine third person plural subject personal pronoun, *i’* serving for both genders (cf. Fr. *ils* masc., *elles* fem.);
- iv. adjectives of color almost invariably precede the noun, e.g., *lé nièr cat* ‘the black cat’ (Fr. *le chat noir*);
- v. the preterite tense, which has gone out of use in informal spoken French, though it survives in the formal written language, is widely used in spoken and written Jèrriais, e.g., *j’donnis* ‘I gave’. Certain third person plural preterite forms are restricted to St. Ouen, e.g., *i’vidrent* ‘they saw’, *i’fùdrent* ‘they went’ (for *i’vîtent*, *i’fûtent* elsewhere)

vi. the imperfect subjunctive, now virtually defunct in spoken French, also survives, e.g., *j'avouais qu'i'l'sûsse* 'I wanted him to know'.

The dominance of English on Jersey is also leading to an increase in frequency of syntactic constructions more isomorphic with English.

Language Planning

Since being introduced into the education system in 1999, Jèrriais is now offered on an extra-curricular basis in most of the Island's schools. It also features in a couple of weekly radio slots and in a fortnightly newspaper column. However, Jèrriais enjoys no more than a token presence on television. Language planning measures receive very little official backing, and with the exception of the Jèrriais education initiative, which has received funding from the States of Jersey, have been left in the hands of groups of enthusiasts.

Jèrriais has been codified via a dictionary (Le Maistre, 1966) and grammar (Birt, 1985). The standard variety is based largely on the sub-dialect of St. Ouen.

Literary Tradition

Although the important 12th-century writer Wace (c. 1100–1179) is known to have come from Jersey, no literary writings in Jèrriais exist until the 19th century. The first author to use the dialect as a medium for his work was Matthieu Le Geyt (1777–1849). Jersey produced several poets and writers during the course of the 19th and 20th centuries; much of their output was published in newspapers and periodicals. Although collections of satirical short stories in Jèrriais were published in pamphlet form from the third quarter of the 19th century onward, the first complete volume of prose to be published in Jèrriais was George Le Feuvre's *Jèrri jadis* (1973).

Bibliography

- Birt P (1985). *Lé Jèrriais pour tous. A complete course on the Jersey language*. Jersey: Don Balleine.
- Brasseur P (1978a). 'Les principales caractéristiques phonétiques des parlers normands de Jersey, Sercq, Guernesey et Magneville (canton de Bricquebec, Manche), Première partie.' *Annales de Normandie* 25(1), 49–64.
- Brasseur P (1978b). 'Les principales caractéristiques phonétiques des parlers normands de Jersey, Sercq, Guernesey et Magneville (canton de Bricquebec, Manche), Deuxième partie.' *Annales de Normandie* 25(3), 275–306.

- Carré A L (1972). *English-Jersey language vocabulary*. Jersey: Don Balleine.
- Falle P (1734). *An account of the island of Jersey* (2nd edn.). London: John Newton.
- Hublart C (1979). 'Le français de Jersey.' Unpublished thesis, Université de l'Etat à Mons.
- Jones M C (2000a). 'Swimming against the tide: language planning on Jersey.' *Language Problems and Language Planning* 24(2), 167–196.
- Jones M C (2000b). 'Ambiguity and unpredictability: linguistic change in modern Jèrriais.' *Verbum* 22(2), 203–222.
- Jones M C (2001). *Jersey Norman French: a linguistic study of an obsolescent dialect*. Oxford: Blackwell.
- Jones M C (2005a). 'Transfer and changing linguistic norms in Jersey Norman French.' *Bilingualism: Language and Cognition* 8(2), 1–17.
- Jones M C (2005b). 'Some structural and social correlates of single word intrasentential codeswitching in Jersey Norman French.' *Journal of French Language Studies* 15 (1), 1–23.
- Joret C (1883). *Des caractères et de l'extension du patois normand*. Paris: Vieweg.
- Lebarbenchon R J (n.d.). *La Grève de Lecq. Littératures et cultures populaires de Normandie 1: Guernesey et Jersey*. Cherbourg: Isoète.
- Le Feuvre G F (1973). *Jèrri jadis*. Jersey: Don Balleine. [2nd edn. 1983.]
- Le Feuvre G F (1976). *Histouaithes et gens d'Jèrri*. Jersey: Don Balleine.
- Le Maistre F (1947). 'The Jersey language in its present state. The passing of a Norman heritage.' Paper read before the Jersey Society in London (July 1947). London: The Jersey Society.
- Le Maistre F (1966). *Dictionnaire Jersiais-Français*. Jersey: Don Balleine.
- Le Maistre F (1979) (1993). *The Jersey language*. (5 cassettes/booklets). Jersey: Don Balleine.
- Liddicoat A J (1994). *A grammar of the Norman French of the Channel Islands: the dialects of Jersey and Sark*. Berlin: Mouton de Gruyter.
- Mason I (1980). 'L'usage des temps narratifs en jersiaise et val de sairais.' *Parlers et Traditions Populaires de Normandie* 47–48, 99–106.
- Scott Warren A (1999). *Lé neu c'mîn*. Jersey: Don Balleine.
- Société Jersiaise (2000). *Les preunmié mille mots*. Jersey: Société Jersiaise.
- Spence N C W (1957a). 'Jèrriais and the dialects of the Norman mainland.' *Bulletin de la Société Jersiaise* 17, 81–90.
- Spence N C W (1957b). 'L'Assibilation de l'r intervocalique dans les parlers jersiais.' *Revue de Linguistique Romane* 21, 270–288.
- Spence N C W (1960). *Glossary of Jersey French*. Oxford: Blackwell.
- Spence N C W (1985). 'Phonologie descriptive des parlers jersiais: I. Les voyelles.' *Revue de Linguistique Romane* 49, 151–165.

Spence N C W (1987). 'Phonologie descriptive des parlers jersiais: II. Les consonnes.' *Revue de Linguistique Romane* 51, 119–133.

Spence N C W (1993). *A brief history of Jèrriais*. Jersey: Don Balleine.

Syvret M & Stevens J (1998). *Balleine's history of Jersey*. West Sussex: Phillimore & Co.

Tapley J (1998). *Jersey Norman French is fun*. Jersey: Don Balleine.

Jewish Languages

B Spolsky, Bar-Ilan University, Ramat Gan, Israel

S B Benor, Hebrew Union College – Jewish Institute of Religion, Los Angeles, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Ethnic and religious groups use language as one means of constructing and expressing their distinctness from other groups. Jews are no exception. Wherever Jews have lived – from Baghdad to Brooklyn, Amsterdam to Odessa – they have spoken somewhat differently from their non-Jewish neighbors. These differences have been as small as the addition of a few Hebrew words and as large as a vastly different lexicon, syntax, and phonology. Therefore, the term “Jewish language” refers to any linguistic variety spoken by Jews that differs to some extent from the non-Jewish language(s) around it. The field of Jewish language studies examines the distinct linguistic practices of the Jewish people around the world.

Jewish languages generally exist in a situation of triglossia with the local non-Jewish language(s) and with a liturgical combination of Hebrew and Aramaic (Weinreich, 1980; Rabin, 1981; and Fishman, 1985). The Jewish language is used mostly for intra-community speech and sometimes for writing. Speakers also generally have at least some knowledge of the co-territorial non-Jewish languages and use them in their interactions with non-Jews. Hebrew and Aramaic have played a very important role in Jewish life. Biblical and rabbinic literatures are studied regularly in their original languages, and daily prayers are conducted mostly in Hebrew and Aramaic. Hebrew is also used for contemporary rabbinic and liturgical production, as well as some other literary functions.

Jewish languages have been documented in many parts of the Jewish diaspora: Yiddish (sometimes referred to as Judeo-German), Judeo-Spanish (also called Ladino, Judezmo, Dzhudezmo, Jidyó, Spanyol, Spanyolit), Judeo-Greek (Yevanic, Romaniyot), Judeo-Italian (Italkic, including local varieties like

Judeo-Venetian), Judeo-Portuguese, Judeo-French (Zarphatic, Western Loez), and Judeo-Provençal (Shuadit) in Europe; Judeo-Arabic (Yahudic), Judeo-Aramaic (Targum, Kurdit), Judeo-Persian (Jidi, Parsic, Judeo-Tadjik, Judeo-Tat, Bukharan), Judeo-Georgian (Gurjuc, Gruzinic), Judeo-Crimean Tatar (Krimchak), and Judeo-Berber in the Middle East, North Africa, and the former Soviet Union; Judeo-Malayalam in India; and Jewish English (Yinglish, Yeshivish) in the New World. Some of the larger and better-studied cases are described in separate sections below.

Scholarship

The scholarly recognition of a phenomenon of Jewish languages goes back to the beginning of the 20th century, when Yiddish became the object of serious academic study. Mises (1915) presented the first large-scale exploration of Jewish linguistic varieties. In late 1970s and early 1980s, Jewish languages started to be studied intensively, following the publication of two major studies of Yiddish that discussed them in a historical context (Birnbaum, 1979; Weinreich, 1980). Around this time, scholars in Israel and the United States edited symposia (Rabin *et al.*, 1979), collections of articles (Paper, 1978; Fishman, 1985; Gold, 1989) and a short-lived journal (Gold and Prager, 1981–1987). More recently, there has been a wave of renewed interest in the subject, as evidenced by the Jewish Language Research Website (www.jewish-languages.org) and the Jewish Languages Mailing List (www.jewish-languages.org/ml).

History

The presumed monolingualism of the early kingdoms of Israel and Judah gave way, in the centuries after the Babylonian exile in the 6th century B.C., to a Hebrew-Aramaic bilingualism (Chomsky, 1957). By the end of the Temple period 2000 years ago, these languages were supplemented by a widespread knowledge of Greek, which was used with distinctive

Jewish features (Wexler, 1985). Thus, Judeo-Aramaic and Judeo-Greek were the earliest Jewish languages that existed in a diglossic relationship with Hebrew.

Judeo-Aramaic was a Jewish adaptation of the major language of wider communication of the Middle East in the millennium before the Common Era. It grew into an important spoken and written Jewish language in Palestine and in the Jewish Diaspora in Babylon, where it was the main language used in the Babylonian Talmud (Greenfield, 1978; Katz, 1985). Among Jews as well as other inhabitants of the region, it was generally replaced by Arabic as a spoken language as a result of the spread of Islam, but it has continued to the present day as a Jewish language in more isolated regions such as Azerbaijan (Garbell, 1965) and Kurdish Iraq (Sabar, 2002).

The third partner in Palestinian trilingualism was Judeo-Greek, widely adapted in Hellenic colonies in Palestine and used by Diaspora Jews throughout the eastern Mediterranean and later in Italy. Judeo-Greek, also called Yevanic, was replaced in most areas starting in the 4th century. An exception is the communities of Romaniote Jews in Greece, which used Judeo-Greek until the influx of Sephardic Jews in the 16th century, when Judeo-Spanish became the majority language of Jews in Greece. Pockets of Judeo-Greek speakers maintained their language in Ioannina, Chalkida, and elsewhere until they were destroyed in the Nazi Holocaust. Few speakers survive today.

Soon after the Roman destruction of Jewish political independence in Palestine in the second century A.D., Hebrew lost its vitality. But it remained firmly entrenched as the language of Jewish religion and literacy, its transmission supported by a religious educational system. Over the next centuries, Jews in exile picked up local languages and developed their own distinctly Jewish varieties, depending in large measure on the nature of their relations with non-Jewish neighbors. As Jews migrated, they generally lost their former language and adapted linguistically to their new land, incorporating distinctive linguistic features. However, two languages defied this trend: Yiddish and Judeo-Spanish, Jewish varieties of Germanic and Hispanic languages, respectively. These languages continued to be used even centuries after their speakers migrated to new lands, where Slavic and Balkan languages were the norm.

In the modern period, when Jews have been able to integrate more fully into some societies, the distinctness of their languages has generally diminished. Yiddish, Judeo-Spanish, and other Jewish languages with long histories have lost significant numbers of speakers due to the combined effects of the Nazi

Holocaust and Jews' cultural and linguistic assimilation into new societies in North America, Europe, and Israel. In the 21st century, Jews generally have full competence in the local languages, vernacular and standard. But their speech also tends to maintain some distinctive features, influenced by Hebrew and Aramaic as well as by the Jewish languages spoken by their ancestors.

Common Linguistic Features

Jewish languages tend to have a number of features in common. Structurally, they are generally based on a spoken variety of a non-Jewish language (Yiddish was based on medieval German and Judeo-Spanish on 15th century Spanish), with a large proportion of borrowings from Hebrew and Aramaic, from earlier Jewish languages, and from other contact languages (Weinreich, 1980). In addition, contemporary Jewish languages tend to be influenced by Israeli Hebrew as a result of affiliations with the State of Israel (Benor, 2004).

The Hebrew and Aramaic influences on Jewish languages are mostly lexical, but some phonological and morphosyntactic influences have been documented as well. Hebrew and Aramaic loan words are most common in the semantic fields of religious life, names of individuals and groups, and euphemism. Until recently, Jewish languages were generally written in Hebrew characters, because of common educational and literacy practices. Orthographic practices have varied, especially in the representation of vowels.

Jewish languages are often strongly influenced by a language spoken by the group's ancestors. In the case of Yiddish, the main previous Jewish language was Judeo-French. In the case of Judeo-Spanish, the main previous Jewish language was Judeo-Arabic. And in the case of Jewish English, the main previous Jewish language was Yiddish. These previous languages provide influences in lexicon, as well as other areas. In addition, the previous languages have a major impact on the use of Hebrew and Aramaic: which words are used, how they are pronounced, and how they are integrated morpho-syntactically.

Most Jewish communities have used the local language in distinctive ways in their translations of biblical and liturgical texts. These translations tend to render the local lexicon in word-for-word imitations of the Hebrew syntax. This practice is referred to in various ways, e.g., Judeo-Arabic *Sharḥ*, Yiddish *Taytsh*, and Judeo-Spanish *Ladino*.

The revitalization and re-vernacularization of Hebrew as part of the Zionist enterprise have produced a new situation, where modern Israeli Hebrew

is markedly distinct from its earlier forms. Is this new variety to be considered a 'Jewish language'? Some have argued that it is too different from Diaspora Jewish languages to be classified with them. On the other hand, it shares many features: strong influence of the previous Jewish language (the Yiddish base is most evident in the highly modified grammar), a special place for Hebrew-Aramaic lexical items, borrowing from the co-territorial non-Jewish languages (including spoken Arabic and the widely known English), and Hebrew orthography.

The Most Widely Spoken Jewish Language: Yiddish

Distinctive linguistic features can be seen in the history of Yiddish, the most widely spoken Jewish language (Birnbaum, 1979; Weinreich, 1980; Katz, 1985; Weigel, 2002). According to the commonly accepted view (Weinreich, 1980), Yiddish was born towards the end of the first millennium A.D. when Judeo-French-speaking Jews started to settle in the Rhineland. During the more tolerant period that preceded the Crusades, these communities shifted from Judeo-French to a variety based on the German spoken in the area. This Judeo-German included elements of Hebrew and Aramaic, as well as other distinctive features. As a result of expulsion, persecution, and changing economic opportunity, many Jews migrated from Germanic-speaking to Slavic-speaking areas and brought their German language with them. In the changed social conditions, the developing Yiddish language maintained its German base while admitting influences from local Slavic languages in lexicon, morphosyntax, phonology, and discourse. In addition, Hebrew and Aramaic elements survived from previous generations, and new ones were added, mostly through contact with liturgical and rabbinic texts. A few Judeo-French lexical elements endured.

Over the centuries, Western Yiddish disappeared as a spoken language, assimilating towards co-territorial German, except in a few areas, like French-speaking Alsace. In eastern Europe, Yiddish developed into a complex web of dialects, differing mostly in phonology but also in lexicon and grammar.

Yiddish documents have been identified as early as the 13th century, and we have examples of epic poems written in Yiddish from the time of the Renaissance. In the early modern period, Yiddish was used mostly in women's religious literature, including translations and explanations of the Bible and liturgy. The mid-19th century saw the flowering of Yiddish literature, stemming from the eastern European Jewish Enlightenment. In the early 20th century, Yiddish became the object of language planning

efforts, including a standardized orthography and linguistic documentation and research.

Events of the 20th century, especially immigration to America and Israel and the Nazi Holocaust, led to a major decline in the use of Yiddish, and today it is used as an everyday language mostly by the elderly and by pockets of Hasidic Jews in the New York area, Israel, and elsewhere (current estimates of total number of speakers range from 200 000 to 400 000). Reversing language shift efforts continue in educational and cultural programs, especially in New York, Montreal, Antwerp, and Mexico City. Young non-Hasidic Jews there and elsewhere continue to use Yiddish as an everyday language in an effort at revitalization.

Judeo-Spanish

Judeo-Spanish is an Hispanic language taken by exiles from Spain after the expulsion of 1492 to northern Europe, the Balkans and Turkey (Sefiha, 1979; Malinowski, 1982; Bunis, 1993; Harris, 1994; Quintana, 2002). There has been much debate about the name of this language: in addition to Judeo-Spanish, commonly used glottonyms are Ladino, Judezmo, and Spanyol, with some scholars maintaining that Ladino should refer only to the calque (word-for-word) translation language variety.

Already in Spain, Judeo-Spanish exhibited influences from Jewish and non-Jewish varieties of Arabic, as well as other distinctive features. When Sephardic Jews migrated, elements of Turkish, Greek, Bulgarian, and other languages were added. In addition, archaisms and independent developments distinguished Judeo-Spanish from contemporaneous peninsular Castilian. Distinctive dialects of Judeo-Spanish formed throughout the Ottoman Empire. In the 19th and 20th centuries, the high-status French language had a major impact, due to influences of religious and secular education.

Judeo-Spanish developed literary functions, including a significant religious literature, a strong oral folk literature, and a corpus of modern belles lettres. There was rapid language loss in the 19th century as a result of emigration, Westernization, and assimilation. The Judeo-Spanish-speaking community in Greece and other Balkan countries was mostly wiped out in the Holocaust. Speakers in Turkey shifted first to French and more recently to Turkish. Today, it is estimated that there are only 30 000–50 000 speakers, mostly elderly.

A North African variety of Judeo-Spanish, called Haketiya, developed in northern Morocco after the 1492 expulsion. Its speakers mostly shifted to Spanish with the establishment of the Spanish Protectorate at the beginning of the 20th century.

Judeo-Arabic

Since even before the Muslim conquest of the Arabian peninsula in the 7th century A.D., Jews have lived alongside Arabic speakers, and they have spoken Jewish varieties of Arabic (Blanc, 1964; Blau, 1981; Hary, 1992; Bar-Asher, 1998). These have included Hebrew and Aramaic influences – mostly lexical, but also phonological, morphological, and syntactic. They have also included archaisms, standardized hyper- and hypo-corrections, and other distinctive features. Judeo-Arabic varieties have been documented in Iraq, Egypt, Syria, Morocco, and Yemen. Due to the migration of Jews within the Arab world, some varieties have features in common with other varieties of Judeo-Arabic that do not exist in the local non-Jewish Arabic dialects.

In the Middle Ages, many important Jewish religious and philosophical works were written in Middle Arabic and Judeo-Arabic. A word-for-word, or calque, translation variety, called *Sharh*, was used for translations of biblical, rabbinic, and liturgical texts. In addition, Judeo-Arabic was used for religious and secular literary production in the 19th century.

Most Jews in Arab lands immigrated to Israel in the 20th century, acquiring Israeli Hebrew and relegating Judeo-Arabic to private use. Those who stayed in Morocco tended to shift to French, and those who immigrated to North America and France tended to shift to the local languages there. It is estimated that there are currently 400 000–500 000 speakers (Grimes, 1996), mostly middle aged and older.

A Contemporary Jewish Language: Jewish English

Also referred to as Judeo-English, Yinglish, and Yes-hivish, Jewish English is an umbrella term for the contemporary in-group varieties spoken by Jews in America, England, and other English-speaking countries (Gold, 1985; Steinmetz, 1986; Weiser, 1994; Benor, 2004). Jewish English is based on the local variety of English with many influences from Yiddish, textual Hebrew and Aramaic, and Israeli Hebrew in lexicon, syntax, phonology, and discourse. Because of the widespread literacy in contemporary English-speaking countries, Jewish English is not written in Hebrew characters. However, Hebrew loan words are sometimes inserted in their original orthography.

The varieties of Jewish English spoken by Orthodox Jews, especially those in larger, more isolated communities, are most distinct from general English, often to the point of being unintelligible to non-Jews

and to non-Orthodox Jews. Orthodox Jewish English is a young language variety, and the number of speakers is growing as the Orthodox community expands. Although Jewish English is the youngest Jewish language that has been researched, it is likely not the only one that is gaining, rather than losing, speakers. As researchers explore the language of other contemporary Diaspora communities, it is expected that they will find similar distinctively Jewish linguistic practices.

Bibliography

- Benor S B (2004). 'Second style acquisition: the linguistic socialization of newly Orthodox Jews.' Ph.D. diss., Stanford University.
- Bar-Asher M (1998). *Linguistic traditions of North African Jews*. Jerusalem: Hebrew University. [In Hebrew].
- Birnbaum S A (1979). *Yiddish: a survey and a grammar*. Toronto: University of Toronto Press.
- Blanc H (1964). *Communal dialects in Baghdad*. Cambridge, MA: Harvard University Press.
- Blau J (1965). *The emergence and linguistic background of Judeo-Arabic: a study of the origins of Middle Arabic*. London: Oxford University Press.
- Blau J (1981). *The emergence and linguistic background of Judeo-Arabic*. Jerusalem: Ben-Zvi Institute.
- Bogoch B (1999). 'Gender, literacy and religiosity: dimensions of Yiddish education in Israeli government-supported schools.' *International Journal of the Sociology of Language* 138, 123–160.
- Bunis D M (1993). *A lexicon of the Hebrew and Aramaic elements in modern Judezmo*. Jerusalem: Magnes Press & Misgav Yerushalayim.
- Chomsky W (1957). *Hebrew: the eternal language*. Philadelphia: The Jewish Publication Society of America.
- Fellman J (1973). *The revival of a classical tongue: Eliezer ben Yehuda and the modern Hebrew language*. The Hague: Mouton.
- Fishman J A (1980). 'Attracting a following to high-culture functions for a language of everyday life: the role of the Tshernovits (Conference in the 'Rise of Yiddish').' *International Journal of the Sociology of Language* 24, 43–73.
- Fishman J A (ed.) (1985). *Readings in the sociology of Jewish languages*. Leiden: E. J. Brill.
- Garbell I (1965). *The Jewish Neo-Aramaic dialect of Persian Azerbaijan*. The Hague: Mouton.
- Gold D (1985). 'Jewish English.' In Fishman J A (ed.) *Readings in the sociology of Jewish languages*. Leiden: Brill. 280–298.
- Gold D (1989). *Jewish linguistic studies*. Haifa: Association for the Study of Jewish Languages.
- Gold D & Prager L (eds.) (1981–1987). *Jewish language review*. Haifa: Association for the Study of Jewish Languages.
- Greenfield J C (1978). 'Aramaic and its dialects.' In Paper H H (ed.) *Jewish languages: themes and*

- variations*. Cambridge, MA: Association for Jewish Studies. 29–43.
- Grimes B A (ed.) (1996). *Ethnologue: languages of the world* (13th edn.). Dallas, Texas: Summer Institute of Linguistics.
- Harris T K (1994). *Death of a language: the history of Judeo-Spanish*. Newark: University of Delaware Press.
- Harshav B (1993). *Language in time of revolution*. Berkeley: University of California Press.
- Hary B (1992). *Multiglossia in Judeo-Arabic: with an edition, translation and grammatical study of the Cairene Purim scroll*. Leiden: Brill.
- Isaacs M (1999). 'Contentious partners: Yiddish and Hebrew in Haredi Israel.' *International Journal of the Sociology of Language* 138, 101–121.
- Jewish Language Research Website: <http://www.jewish-languages.org/>.
- Katz D (1985). 'Hebrew, Aramaic and the rise of Yiddish.' In Fishman J A (ed.) *Readings in the sociology of Jewish languages*. Leiden: E. J. Brill. 85–103.
- Malinowski A (1982). 'A report on the status of Judeo-Spanish in Turkey.' *International Journal of the Sociology of Language* 37, 7–23.
- Mieses M (1915). *Die Entstehungsursache der jüdischen Dialekte*. Vienna: R. Löwit Verlag.
- Paper H H (ed.) (1978). *Jewish languages: themes and variations*. Cambridge, MA: Association for Jewish Studies.
- Pilowsky A L (1985). 'Yiddish alongside the revival of Hebrew: public polemics on the status of Yiddish in Eretz Israel, 1907–1929.' In Fishman J A (ed.) *Readings in the sociology of Jewish Languages*. Leiden: Brill. 104–124.
- Quintana A (2002). 'Geografía lingüística del Judeo-español de acuerdo con el léxico.' *Revista de Filología Española* 82(1–2), 105–138.
- Rabin C (1981). 'What constitutes a Jewish language?' *International Journal of the Sociology of Language* 30, 19–28.
- Rabin C *et al.* (1979). 'The Jewish languages: commonalities, differences and problems (in Hebrew).' *Pe'amim* 1, 40–57.
- Sabar Y (2002). *A Jewish Neo-Aramaic dictionary: dialects of Amidya, Dibok, Nerwa and Zakho, northwestern Iraq*. Wiesbaden: Harrassowitz.
- Sephiha H V (1979). *Le Ladino (Judeo-Espagnol calque)*. Paris: Association Vidas Largas.
- Spolsky B (1991). 'Hebrew language revitalization within a general theory of second language learning.' In Cooper R L & Spolsky B (eds.) *The influence of language on culture and thought: essays in honor of Joshua A. Fishman's sixty-fifth birthday*. Berlin: Mouton de Gruyter. 137–155.
- Spolsky B & Shohamy E (1999). *The languages of Israel: policy, ideology and practice*. Clevedon: Multilingual Matters.
- Steinmetz S (1987). *Yiddish and English: a century of Yiddish in America*. University: University of Alabama Press.
- Weigel W (2002). 'Yiddish.' *Jewish language research website*. <http://www.jewish-languages.org/yiddish.html>.
- Weinreich M (1980). *History of the Yiddish language*. Noble S & Fishman J (trans.). Chicago: University of Chicago Press.
- Weiser C (1995). *FrumSpeak: the first dictionary of Yeshivish*. Northvale, NJ: Jason Aronson.
- Wexler P (1981). 'Jewish interlinguistics: facts and conceptual framework.' *Language* 57, 99–149.
- Wexler P (1985). 'Recovering the dialects and sociology of Judeo-Greek in non-Hellenic Europe.' In Fishman J A (ed.) *Readings in the sociology of Jewish languages*. Leiden: E. J. Brill. 227–240.

Relevant Website

<http://www.jewish-languages.org> – The Jewish Language Research Website.

Jiwarli

P K Austin, School of Oriental and African Studies, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Jiwarli (Djiwarli) is an Australian Aboriginal language and was traditionally spoken along the upper reaches of the Henry River, a tributary of the Ashburton River, in the northwest of Western Australia. The language was unrecorded until 1978 and is now extinct, the last speaker, Mr. Jack Butler, having passed

away in May 1985. Before his death, Jack Butler worked with Peter Austin to record over 70 texts in a range of genres, a lexicon of some 1500 words and elicitation of morphological paradigms and syntactic constructions. Publications on the language include a bilingual dictionary (Austin, 1992), a text collection (Austin, 1997), articles on morphosyntax (Austin and Bresnan, 1996; Austin, 1995, 1998, 2000, 2001) and a website. The language has become known in the linguistic literature for its nonconfigurational syntax (see Austin and Bresnan, 1996; Baker, 2000, and below), and it also shows switch-reference and a complex system of case-marking that reflects

clause type (Austin, 2004). A reference grammar is in preparation.

Language Relationships

Jiwarli is closely related to its immediate neighbors, Warriyangka (Warriyangga), Thiin, and Tharrkari (Dhargari) as members of the Mantharta subgroup (*mantharta* being the word for ‘person’). The languages share up to 80% common vocabulary and a similar grammatical system. Tharrkari has undergone a number of historical phonological changes that make its phonetics and phonology highly unusual for an Australian language (see further below). None of the Mantharta languages has any native speakers today, though some knowledge of words and expressions remains among descendants. The Mantharta languages are most closely related to the Kanyara languages spoken to their west and northwest: Payungu (Bayungu), Pinikura (Binigura), Purduna (Burduna), and Thalanyji (Dhalandji). They share approximately 60% cognate vocabulary and have a number of grammatical features in common, including switch-reference and clause linkage effects on case-marking (Austin, 1996, 2004). Today only Thalanyji continues to be spoken by older members of a single family living near Onslow, Western Australia. The Kanyara and Mantharta languages belong to the widespread Pama–Nyungan family, which covers the southern two-thirds of Australia (see **Australia: Language Situation**), and are most closely related to the Nyungic languages spoken to their east.

Linguistic Characteristics

Phonology

The phonological system of Jiwarli is typical of languages of the region, with contrastive stops at six points of articulation, a nasal for each stop position, a lateral for each nonperipheral stop, a flap, a semi-retroflex continuant, and two glides. **Table 1** gives the relevant consonants in their practical orthographic form. There are just three vowels: high front *i*,

high back *u*, and low *a*, with a phonemic length contrast mainly, but not exclusively, found in the first syllable of words. Tharrkari has undergone a number of historical phonological changes that have resulted in the creation of a stop voicing contrast (unusual in Australia) and, in one dialect, complete loss of laterals.

The general structure of Jiwarli roots is CV(C)CV(C). Every word in Jiwarli must begin with a consonant and end in a vowel; roots can end in a consonant but if otherwise unsuffixed *-ma* is added to nasal-final roots and *-pa* to roots ending in *l*, *rl*, *ly*, or *rr*. Word-initially only nonapico-domal stops and nasals and the two glides are found. Word-medially there are limited consonant clusters, primarily homorganic nasal plus stop, and apical nasal or lateral plus peripheral stop (*p* and *k*). Vowel clusters are not found. Words borrowed from English are generally restructured to meet these phonotactic constraints, e.g., *walypala* ‘white man’ (from ‘white fella’), *ngayirlanma* ‘island.’

Morphology

Jiwarli, like other languages of the Pama–Nyungan group, is entirely suffixing in its morphology. There are two major word classes, nominals and verbs, with nominals showing a rich system of case-marking and verbs marking tense/aspect/mood and dependent clause categories. Nominals can be subdivided into substantives (which cover both noun and adjective concepts in a language like English), pronouns, locationals, and demonstratives. Minor word classes include adverbs, particles, and interjections.

Nominals in Jiwarli inflect for case, with the syntactic functions of intransitive subject (S), transitive subject (A), and transitive object (P) showing a split-ergative pattern of syncretism in the case forms determined by animacy:

- for the first person singular pronoun, S and A fall together as a single (unmarked) form
- for inanimate nominals and demonstratives, S and P fall together as a single (unmarked) form

Table 1 Consonants

| | Bilabial | Lamino- | | Apico- | | Dorsal-velar |
|------------|----------|-----------|-----------|-----------|-----------|--------------|
| | | dental | palatal | alveolar | domal | |
| Stop | <i>p</i> | <i>th</i> | <i>j</i> | <i>t</i> | <i>rt</i> | <i>k</i> |
| Nasal | <i>m</i> | <i>nh</i> | <i>ny</i> | <i>n</i> | <i>rn</i> | <i>ng</i> |
| Lateral | | <i>lh</i> | <i>ly</i> | <i>l</i> | <i>rl</i> | |
| Flap | | | | <i>rr</i> | | |
| Continuant | | | | | <i>r</i> | |
| Glide | <i>w</i> | | <i>y</i> | | | |

- for all other nominals, there are three forms for S, A, and P functions

In addition to the three main cases (nominative for S, ergative for A, accusative for P) there are also the following case forms:

- dative, marking alienable possession and complement of certain verbs
- allative, coding direction toward a place
- locative, coding location in a place, and complement of verbs of speaking
- ablative, coding direction from a place and cause

The actual forms of the cases are affected by the phonological shape of the root, e.g., whether it ends in a vowel or not, what kind of vowel or consonant is root-final, and how many morae it contains (long vowels counting as two mora). **Table 2** sets out a sample substantive declension.

The coding of transitive object P varies according to clause type and crossclausal reference relations: in certain dependent clauses (for details, see below) P is marked as dative or as allative. In addition, case is added to dependent clause verbs to indicate cross-clausal coreference (see below), and manner

adverbs in transitive clauses carry an ergative case agreement marker.

Jiwarli has a rich system of nominal word-building morphology that involves suffixation between the root and case inflection. Categories encoded in word-building morphology include number (dual, paucal, plural), having (e.g., *yakan-jaka* ‘married [lit. spouse-having]’), lacking (e.g., *yakan-yirra* ‘unmarried [lit. spouse-lacking]’), and kin dual and plural (e.g., *kurta* ‘older brother,’ *kurtarra* ‘pair of brothers’).

Pronouns in Jiwarli distinguish three persons and singular, dual, and plural number; in the first person non-singular there is an inclusive–exclusive contrast. Demonstratives encode a proximal and distal contrast. **Table 3** sets out the basic pronoun and demonstrative forms.

Verbs morphologically distinguish between main verb and dependent verb inflections. Main verbs encode tense/aspect/mood categories such as past habitual, present, future, and imperative. Dependent verbs occur in hypotactically linked clauses and encode clause type (relative tense plus aspect) plus cross-clausal coreference or non-coreference of

Table 2 Substantive cases

| | A | S | P | Dative | Locative | Allative |
|-----------|--------------------|-------------------|---------------------|-------------------|---------------------|----------------------|
| boy | <i>wirtangku</i> | <i>wirta</i> | <i>wirtanha</i> | <i>wirtawu</i> | <i>wirtangka</i> | <i>wirtarla</i> |
| girl | <i>kurlkingku</i> | <i>kurlki</i> | <i>kurlkinha</i> | <i>kirkiyi</i> | <i>kurlkingka</i> | <i>kurlkirla</i> |
| dog | <i>thuthungku</i> | <i>thuthu</i> | <i>thuthunha</i> | <i>thuthuwu</i> | <i>thuthungka</i> | <i>thuthurla</i> |
| fire | <i>karlangku</i> | <i>karla</i> | <i>karla</i> | <i>karlawu</i> | <i>karlangka</i> | <i>karlarla</i> |
| tree | <i>wurungku</i> | <i>wuru</i> | <i>wuru</i> | <i>wuruwu</i> | <i>wurungka</i> | <i>wururla</i> |
| hill 'roo | <i>mathantu</i> | <i>mathanma</i> | <i>mathannha</i> | <i>mathanku</i> | <i>mathanta</i> | <i>mathankurla</i> |
| tongue | <i>thalanythu</i> | <i>thalanyma</i> | <i>thalanyma</i> | <i>thalanyku</i> | <i>thalanytha</i> | <i>thalanykurla</i> |
| chin | <i>nyinyarntu</i> | <i>nyinyarnma</i> | <i>nyinyarnma</i> | <i>nyinyarnku</i> | <i>nyinyarnta</i> | <i>nyinyarnkurla</i> |
| wind | <i>yuwalpalu</i> | <i>yuwalpa</i> | <i>yuwalpa</i> | <i>yuwalku</i> | <i>yuwalpala</i> | <i>yuwalkurla</i> |
| cousin | <i>ngathalpalu</i> | <i>ngathalpa</i> | <i>ngathalpanha</i> | <i>ngathalku</i> | <i>ngathalpala</i> | <i>ngathalkurla</i> |
| barb | <i>ngarlirpalu</i> | <i>ngarlirpa</i> | <i>ngarlirpa</i> | <i>ngarlirru</i> | <i>ngarlirrpala</i> | <i>ngarlirrkurla</i> |

Table 3 Pronouns and demonstratives

| | A | S | P | Dative | Locative |
|---------|-----------------------|---------------------|------------------------|------------------------|-----------------------|
| 1sg | <i>ngatha</i> | <i>ngatha</i> | <i>ngathanha</i> | <i>nganaju</i> | <i>ngathala</i> |
| 1dincl | <i>ngalilu</i> | <i>ngali</i> | <i>ngalinha</i> | <i>ngalimpa</i> | <i>ngalila</i> |
| 1dlexcl | <i>ngalijuru</i> | <i>ngaliju</i> | <i>ngalijunha</i> | <i>ngalijungu</i> | <i>ngalijura</i> |
| 1plincl | <i>nganthurrulu</i> | <i>nganthurru</i> | <i>nganthurranha</i> | <i>nganthurrampa</i> | <i>nganthurrala</i> |
| 1plexcl | <i>nganthurrajuru</i> | <i>nganthurraju</i> | <i>nganthurrajunha</i> | <i>nganthurrajungu</i> | <i>nganthurrajura</i> |
| 2sg | <i>nhuralu</i> | <i>nhurra</i> | <i>nhurranha</i> | <i>nhurrampa</i> | <i>nhurrala</i> |
| 2dl | <i>nhupaluru</i> | <i>nhupalu</i> | <i>nhupalunha</i> | <i>nhupalumpa</i> | <i>nhupalura</i> |
| 2pl | <i>nhurarakaru</i> | <i>nhurakara</i> | <i>nhurakaranha</i> | <i>nhurakarampa</i> | <i>nhurakarala</i> |
| 3sg | <i>panhaluru</i> | <i>panhalu</i> | <i>panhalunha</i> | <i>pannumpa</i> | <i>panhalura</i> |
| 3dl | <i>pulalu</i> | <i>pula</i> | <i>pulanha</i> | <i>pulampa</i> | <i>pulala</i> |
| 3pl | <i>thanalu</i> | <i>thana</i> | <i>thananha</i> | <i>thanampa</i> | <i>thanala</i> |
| this | <i>yilu</i> | <i>yinha</i> | <i>yinha</i> | <i>yimu</i> | <i>yila</i> |
| that | <i>ngulu</i> | <i>ngunha</i> | <i>ngunha</i> | <i>ngurnu</i> | <i>ngula</i> |

Table 4 Verb inflections

| Inflection | Conj 1 | Conj 2 | Conj 3 | Conj 4 | Conj 5 |
|-----------------------------------|----------------|----------------|--------------|-------------|-------------|
| Main clause verb inflections | | | | | |
| Usitative | -laartu | -rraartu | -artu | -artu | -artu |
| Past | -rninyja | -rninyja | -nyja | -nyja | -nyja |
| Present | -nha | -nha | -inha | -inha | -a |
| Future | -lka | -rrka | -ira | -ra | -ra |
| Imperative | -nma | -nma | -ma | -ma | -ma |
| Irrealis | -nmararni | -nmararni | -mararni | -mararni | -mararni |
| Dependent clause verb inflections | | | | | |
| ImperfSS | -rnu | -rnu | -nhu | -nhu | -nhu |
| ImperfDS | -niya | -niya | -iniya | -iniya | -iniya |
| PerfSS | -rninyjalu | -rninyjalu | -nyjalu | -nyjalu | -nyjalu |
| PerfDS | -rninyjaparnti | -rninyjaparnti | -nyjaparnti | -nyjaparnti | -nyjaparnti |
| PurpSS | -ru | -rru | -yi | -ngku | -a |
| PurpDS | -lpuka | -rrpuka | -puka | -puka | -puka |
| Intentive | -lkarri(ngu) | -rrkarri(ngu) | -irarri(ngu) | -rarri(ngu) | -rarri(ngu) |
| Might | -lkangu | -rrkangu | -irangu | -rangu | -rangu |

subjects (S or A), i.e., switch-reference (see further details in the syntax section below). There are five morphologically determined verb conjugations: conjugations one and two are primarily, but not exclusively, transitive, and conjugations three, four, and five are intransitive. Table 4 sets out the verb conjugations.

Verbs show limited word-building morphology, mainly transitivity and detransitivizing affixes that shift conjugation and transitivity. There are also category-changing affixes:

- nominalizing suffixes that create agent or instrument nominals from verbs
- verbalizing suffixes that create intransitive (inchoative) or transitive (causative) verbs from nominals.

The minor categories of adverb, particle, and interjection show no morphological variation. However, there is a set of postinflectional suffixes that may be attached to words of any class to encode various information status concepts, such as *-rru* for ‘new information’ and *-thu* for ‘old information.’ These affixes are ubiquitous in texts.

Syntax

Jiwarli is a nonconfigurational language (Hale, 1983; Austin and Bresnan, 1996; Baker, 2000) and shows the following syntactic characteristics:

- free word order, in which any possible order of sentence constituents is found (Austin, 2000)
- split-NP syntax, in which nominals understood as referring to a single entity can be separated in the clause by other constituents (each nominal bearing a relevant case marker)

- free argument ellipsis, in which nominals of any person or number whose reference is clear from the context can be freely omitted (Austin, 2001).

The following example illustrates split-NP syntax:

- (1) *Karla wanta-nma-rni jarnpa juma*
 fire.ACC give-IMPER-hence light.ACC small.ACC
 ‘Give me a small fire light’ [T52s15]

Free ellipsis of arguments is seen in the following (see also line 68 in the text example below):

- (2) *Wirntupinya-nyja-rru*
 kill-PAST-NEW-INF
 ‘(They) killed (him)’ [T42s25]

Jiwarli also shows interesting interclausal syntax. Dependent clauses occur hypotactically located on the margins of main clauses, and their verbs encode clause type plus switch-reference, i.e., (non-)coreference of subject (S or A) between the main and dependent clause. In same-subject clauses the dependent clause subject is obligatorily unexpressed (these being ‘control’ structures). When the main clause is transitive, some same-subject dependent clause verbs carry an ergative case marker in agreement with the controlling subject nominal. The following examples illustrate this:

- (3) *Mantharta kumpa-inha wurnta-wu*
 man.NOM sit-PRES shield-DAT
yinka-rnu
 adze-IMPERESS
 ‘The man sits adzing a shield’ [N11p31s3]
- (4) *Nhurra-kara-lu thika-nma*
 you-PL-ERG eat-IMPER
yarrukarri-ngu-ru-thu
 want-IMPERESS-ERG-old-INF
 ‘You eat it if you want it!’ [N11p39s3]

Table 5 Coding of P

| | Dependent object |
|-----------------|------------------|
| Intensive | Dative |
| Imperfective-ss | |
| Perfective-ss | |
| Imperfective-ds | |
| Perfective-ds | Allative |
| Purposive-ss | |
| Purposive-ds | Accusative |
| Might | |

For different-subject dependent clauses, if there is coreference between the (omitted) subject of the dependent clauses and a nonsubject in the main clause, an agreement case marker appears on the dependent verb, as in:

- (5) *Tharla-nma yinha julyu-nba*
 feed-IMPER this.ACC old man-ACC
kamu-rri-ya-nba
 hunger-INCHOAT-IMPEREDS-ACC
 ‘Feed this old man who is becoming hungry!’
 [JIT13s1]

Notice that there is a complex interaction between the marking of P inside the dependent clause (as dative, allative, or accusative) depending on clause type and crossclausal coreference relations. Table 5 illustrates this.

The significance of these patterns is explored more generally in Austin (2004).

Particles in Jiwarli have scope over the whole clause and encode such semantic concepts as negation, possibility, etc. An example is *warri* ‘not’ in (see also *kaji* ‘try’ in text line 15 below):

- (6) *Nhaa-rri-nyja nhurra warri*
 what-INCHOAT-PAST you.ERG not
kurlkayi-rnu wangka-iniya-wu nganaju
 listen-IMPERESS talk-IMPEREDS-DAT I.DAT
 ‘Why didn’t you listen to me talking?’ [T35s7]

Text Example

The following extract (Text 43 in Austin, 1997) exemplifies the morphological and syntactic characteristics of Jiwarli and shows a little of the cultural background of the language. It comes from a traditional story in which one bird steals fire from the people, who then ask the Peregrine falcon to get it back from the thief:

- (7) *Ngana-lu ngunha karla mana-ra*
 who-ERG that.ACC fire.ACC get-FUT
 ‘Who will get the fire?’

- (8) *Nhurra parru*
 2SG.NOM then
 ‘How about you?’
- (9) *Nhurra karlathintirnira kurukurura,*
 2SG.NOM Peregrine falcon Peregrine falcon
nhurra yini-thu
 2SG.NOM name.NOM-OLD-INF
 ‘You are karlathintirnira Peregrine falcon, (that’s) your name.’
- (10) *Ngaa*
 yes
 ‘Yes’
- (11) *Ngunha thurni-nyja-nthi*
 that.NOM laugh-PAST-just
 ‘He just laughed’
- (12) *Yana-nyja ngunha purtipala-rru*
 go-PAST that.NOM pretty.NOM-NEW-INF
 ‘He was pretty now’
- (13) *Wantha-rninyja juuri wangkarr-a*
 put-PAST paint.ACC throat-LOC
 ‘(They) put paint on (his) throat’
- (14) *Wantha-rninyja kala-pa wangkarr-a*
 put-PAST like this-SPEC throat-LOC
 ‘(They) put (it) like this on (his) throat’
- (15) *Kaji nhurra yana-ma mana-ngku*
 try 2SG.NOM go-IMPER get-PURP.SS
ngurlu karla-rla
 that.ALL fire-ALL
 ‘You try to go and get the fire’

Bibliography

- Austin P (1992). *A dictionary of Jiwarli, Western Australia*. Melbourne: La Trobe University.
- Austin P (1995). ‘Double case marking in Kanyara and Mantharta languages.’ In Plank F (ed.) *Agreement by Suffixaufnahme*. Oxford: Oxford University Press. 363–379.
- Austin P (1996). ‘Kanyara and Mantharta languages.’ University of Melbourne. [Prepared for Australian Linguistic Institute Workshop, ‘Where did the Western Desert language come from?’ convened by Patrick McConvell (NTU) and Mary Laughren (UQld).]
- Austin P (1997). *Texts in the Mantharta languages, Western Australia*. Tokyo: Tokyo University of Foreign Studies.
- Austin P (1998). ‘Eaglehawk was sitting chasing them: grammaticisation of the verb ‘to sit’ in Mantharta languages, Western Australia.’ In Siewierska A & Jung Song J (eds.) *Case, typology and grammar: in honour of Barry J. Blake*. Typological Studies in Language 38. Amsterdam: John Benjamins.
- Austin P (2000). ‘Word order in a free word order language: the case of Jiwarli.’ In Simpson J, Nash D, Laughren M, Austin P & Alpher B (eds.) *Forty years on: Ken Hale and Australian languages*. Canberra: Pacific Linguistics. 305–324.

- Austin P (2001). 'Zero arguments in Jiwarli, Western Australia.' *Australian Journal of Linguistics* 21(1), 83–98.
- Austin P (2004). 'Case and clauses in Australian Aboriginal languages.' University of London, manuscript.
- Austin P & Bresnan J (1996). 'Non-configurationality in Australian Aboriginal languages.' *Natural Language and Linguistic Theory* 14, 215–268.
- Baker M (2001). 'Non-configurationality.' In Baltin M & Collins C (eds.) *The handbook of syntactic theory*. London: Blackwells.
- Hale K (1983). 'Warlpiri and the grammar of non-configurational languages.' *Natural Language and Linguistic Theory* 1(1), 5–74.

K

Kalkutungu

B J Blake, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

Kalkutungu (also Kalkatungu (Kalkatung) and Kalkadoon) is a Pama-Nyungan language, and, like many other Australian languages, it is largely agglutinative, although with a good deal of irregular bound forms. Nouns are inflected for case and verbs for tense, aspect, mood, and modality. Word order is controlled by discourse principles with few grammatical constraints.

Kalkutungu has a typical phoneme inventory with two laminal series of stops, nasals, and laterals. In company with the Arandic languages to the west, at some stage in the past Kalkutungu underwent a change involving the loss of initial consonants and in some instances initial syllables. This has resulted in words with initial *a*, such as *arnka* 'to be ill' (cf. Yalarnnga *yarnka*) and words with initial homorganic nasal-stop clusters such as *mpaya* 'you two' (cf. Yalarnnga *nhumpala*).

The core case marking opposes an ergative case for the agent of a transitive verb (A) and an unmarked absolutive (alternatively nominative) for an intransitive subject (S) and object (O). The ergative-absolutive distinction applies to free pronouns as well as to nouns, but there is a system of clitic pronouns that oppose a subject set covering S and A functions to an object set. An odd feature of Kalkutungu is that although these clitics are obligatory with auxiliaries and in the perfect and imperfective aspects, they are otherwise optional and used rather sparingly. The following examples illustrate the core case marking and clitic pronouns. Note in (2) that the ergative expresses the instrument used to perform an action as well as the agent.

- (1) marapayi ingka- ntiya- tawun-
 nha-na piyangu kunha
 woman. go-PT- hill-ABL town-
 ABS 3PL.SUBJ ALL
 'the women went from the hill to the town'
- (2) marapayi- ngayi kanimayi- ru-
 thu nha-ngi-na ngku

woman-ERG me.ABS tie-PT-1SING. rope-ERG
 OBJ-3PL.SUBJ
 'the women tied me up with rope'

There is a derived two-place intransitive construction, the antipassive, in which the subject appears in the absolutive and the object is demoted to the dative. In independent clauses, this construction is used to signal reduced semantic transitivity. It is used, for instance, to indicate a generic object as in (3b) rather than a specific object as in (3a), and it is obligatory in the imperfective.

- (3a) marapai-thu rumpa-mi ithirr
 woman-ERG grind-FU seed.ABS
 matyamirla-thu
 grindstone-ERG
 'the woman will grind the seed with the
 grindstone'
- (3b) marapai rumpa-yi-mi
 woman.ABS grind-AP-FU
 ithirr-ku matyamirla-thu
 seed-DAT grindstone-ERG
 'the woman will grind seed with the grindstone'

In dependent clauses, the antipassive is used to signal co-reference between the agent of the dependent clause and the S or O of the governing clause. In (4), (5), and (6), we have a type of subordinate clause used to express purpose adjuncts and complements to directive verbs. These are characterized by a purpose auxiliary, which hosts clitic pronouns. In (4), the agent of the purpose clause is co-referent with the S of the governing clause so the purpose clause appears in the antipassive, as is evident from the marking on the verb and the dative marking for the patient. In (5), the agent of the subordinate clause is co-referent with the O of the governing clause, so the antipassive is used. In (6), however, the agent of the purpose clause is coreferent with the allative-marked *mangarnaan-kunha* 'to the doctor', so no antipassive is used.

- (4) ngayi ingka-nha nha-wu thuwarr-ku
 I.ABS go-PT this-DAT snake-DAT
 lhaa ngulurrma-yi
 PURP.1SING.SUBJ catch-AP
 'I went to catch the snake'

- (5) nga-thu tjaa pati-nha thuku-wu
 I-ERG this.ABS tell-PT dog-DAT
 kuntu a-yi ngayima-yi
 not PURP-3SING.SUBJ chase-AP
 'I told this one not to chase the dog'

- (6) nhaa nga-thu unpiyi nga-tji nhawurr
 this.ABS I-ERG take I-DAT child.ABS
 mangarnaan-kunha a-yi puthurr-puni
 doctor-ALL PURP-3SING.SU good-make
 'I took my child to the doctor for him/her to cure her'

A feature that Kalkutungu has in common with a number of other Australian languages is insubordination, whereby a subordinate clause becomes independent. Sentence (7a) provides an example; compare it with (4), (5), and (6). Essentially, a verb such as *ingka* in (4) is redundant and over time such verbs have been dropped in some circumstances to yield a construction in which the erstwhile dependent clause is independent. Witness the ergative on the agent in (7a), compared to the absolutive in (4). This construction, in which clitic pronouns are obligatory, makes an interesting comparison with (7b), in which the future tense is used and no clitics.

- (7a) nyin-ti a-ngi lha?
 you-ERG PURP-1SING.OBJ hit
 'are you going to hit me?'
- (7b) nhakaakuwa nyin-ti ngayi lhami?
 why you-ERG me.ABS hit-FU
 'why are you going to hit me?'

Kalkutungu has two applicative constructions. In one, a dative can be promoted to object. If there is a patient object, this is retained, so there is a double object construction. In sentence of (8a), the beneficiary is expressed in the dative. In (8b), the beneficiary is now O and is expressed by an object clitic.

- (8a) utjan nga-tji intji-ya
 firewood I-DAT chop-IMP
 'chop some wood for me'

- (8b) utjan intji-tjami-ya-ngi
 firewood chop-APPL-IMP-1SING.OBJ
 'chop me some wood'

In the other applicative, an instrumental or locative can be promoted. If there is a patient object, it is demoted to the dative. The principal function of this construction is to provide for the anaphoric deletion of the instrumental or locative in purpose or other subordinate clauses. Example (9) is typical. The knife is understood to be the instrument in the purpose clause, but it is covert. The patient is demoted to the dative. Note this demotion is not in response to co-reference between the agent of the purpose clause and S or O, as in (4) and (5). Note too that the same marker *-nti* functions as a causative and applicative.

- (9) kankari iti-nti-ya ati-ntji
 knife. ABS return-CAUS-IMP meat-DAT
 lhaa pintji-nti
 PURP-1SING. SUBJ cut-AP
 'bring the knife so I can cut the meat with it'

The final example is another sample of the instrumental applicative in which the instrument *utjula* 'net' has been promoted to O, pushing the patient *wakarri* 'fish' into the dative. It serves to illustrate two common principles of discourse. First, it is common to represent a nominal with a pronoun or other generic expression early in the clause with a more specific noun later. In (10), *nhaa* 'this' anticipates *utjula* 'net'. Second, the focus is normally placed first in the clause, in this instance *yawun* 'big'.

- (10) yawun nhaa nga-thu utjula
 big.ABS this.ABS I-ERG net.ABS
 wakarri-yi ngurlurr-manti
 fish-DAT catch-AP
 'I use a big net to catch fish with'

Bibliography

Blake B J (1979). *A Kalkatungu grammar*. Canberra, Australia: Pacific Linguistics.

Kannada

S N Sridhar

© 2006 Elsevier Ltd. All rights reserved.

Kannada (formerly referred to sometimes as 'Kana-rese' or 'Canarese') is a South Dravidian language spoken by about 28 million people primarily in the state of Karnataka, South India, where it is the

official language. It is also spoken in the neighboring states of Tamil Nadu, Andhra Pradesh, Kerala, and Maharashtra.

History

Kannada is written in a variety of the 'alpha-syllabic' Brahmi system. The earliest written record, the

Halmidi inscription (ca. 450 A.D.), already shows the influence of Sanskrit. The history of the language is conventionally divided into three periods, Old Kannada (up to ca. the thirteenth century A.D.), Middle Kannada (up to ca. the nineteenth century), and Modern Kannada (the nineteenth century onward).

Literary and Grammatical Tradition

Kannada has a world-class literary tradition, but few works have been translated. It also has a rich grammatical tradition. The first extant literary work, Nagavarma's *Kavirajamarga* (ninth century), is a treatise on poetics and includes the earliest grammar of Kannada. Kesiraja's *Śabdamanidarpana* (1260) is the classic grammar of Old Kannada. The other major grammar is Bhattakalanka's *Karna:taka Śabda:nus̥a:-sana* (1604), written in Sanskrit. Modern descriptions include Kittel (1903), Spencer (1950), Bright (1958), Schiffman (1983), and Sridhar (1989). Srikanthaiah (1960) and Bhat (1978) include other important descriptions in Kannada.

Variation

Three major regional varieties are recognized: the dialect of the former and present capitals – the Mysore/Bangalore Kannada, the Dharwar (northern) Kannada, and the Mangalore (coastal) Kannada. The Brahmin variety of Mysore Bangalore forms the basis of the standard language. It retains aspirated consonants and consonant cluster combinations in Sanskrit-derived words (which occur with high frequency in this variety). There are other caste varieties, which differ in phonology, morphology, syntax, and the lexicon. Kannada is a diglossic language.

Kannada is interesting from the point of view of historical sociolinguistics because of the millennium-old controversy between the elitists and populists on the one hand and the purists and the pragmatists on the other concerning the proper literary style – a debate that continues to this day.

Structure

Kannada morphology is agglutinative, suffixing, and quite regular. Nouns are marked by cases, postpositions, number, and occasionally gender. Verbs indicate tense, aspect, and agreement in number, person, and gender. Negation is expressed as main verb and as a suffix. Word order is subject–object–verb. The verb normally (but not always) ends the sentence. Modifiers (e.g., adjectives, adverbs, and subordinate clauses) precede their heads; auxiliaries follow

main verbs. Compounding is very common in nouns and verbs. Normally, only one finite verb occurs per sentence: the syntax relies heavily on participles, gerunds, and infinitives. A number of these points are illustrated in the following sentence:

ni:nu nenne meccida a: mu:ru puṭṭa
you yesterday admired those three little

makkaḷu a:ṭa a:ḍi
children game play – past participle
danidu waḷamaneyalli nidde
tire – past participle inside room – locative sleep
ho:gidda:re.

go – past participle – continuous – 3 (human) plural
'Those three little children whom you admired
yesterday have gone to sleep after playing and
getting tired.'

Importance of Language Contact

Kannada has successfully assimilated an enormous amount of Sanskrit, Prakrit, Perso–Arabic, New Indo–Aryan, and more recently, English elements. This influence is manifested in the rather substantial stock of loan words, the productive use of Sanskrit derivational morphology, code-mixing, and calquing, which have been the preferred strategies for modernization throughout its history. This openness to borrowing sharply distinguishes Kannada from its sister language, Tamil.

Kannada and Linguistic Theory

Evidence from Kannada has played an influential role in several areas of modern linguistic inquiry. These include the status of grammatical relations, especially nonnominative subjects, in syntactic theory; processes of syntactic and morphological convergence; morphological levels in derivation; caste dialects; and syntactic and psycholinguistic models of bilingual code-switching, among others.

Bibliography

- Bhat D N S (1978). *Kannada Vakyagalu: Antarika racane mattu Artha Vyavasthe*. Mysore: Geeta Press.
Bhattakalanka (1604). *Karnataka S'abda: nus'a: sanam*. Bangalore: Raj Kamal Prakashana.
Bright W (1958). *An Outline of Colloquial Kannada*. Poona: Deccan College Postgraduate and Research Institute.
Kesiraja (1260). *S'abdamaniḍarpana* Madras: University of Madras.
Kittel F (1903). *A Grammar of the Kannada language in English, comprising the three dialects of the language*

(*Ancient, Mediaeval and Modern*). Mangalore: Basel Mission Book and Tract Depository.
 Nagavarma (ca. 9th century). *Karnataka Bhasha Bhushana*. New Delhi: Asia Educational Services.
 Schiffman H F (1983). *A reference grammar of spoken Kannada*. Seattle: University of Washington Press.

Spencer H (1950). *A Kanarese grammar*. Mysore: Wesley Press.
 Sridhar S N (1989). *Kannada*. London: Routledge.
 Srikanthaiah T N (1960). *Kannada Madhyama Vya:karana*. Gubbi: Sevasadana Press.

Kanuri

G J Dimmendaal, University of Cologne, Cologne, Germany

© 2006 Elsevier Ltd. All rights reserved.

One of the earliest documents on an African language, a short vocabulary dating back to the 17th century, involves data from Kanuri, also known as Yerwa, or Bornu; the name 'Beriberi' is considered to be derogatory by Kanuri speakers. Kanuri is a major language in terms of the number of speakers; estimates for Nigeria range between three and four million, whereas in neighboring countries like Cameroon, Chad, and Niger, there may be around half a million speakers. Kanuri forms a dialect cluster with Kanembu, which is spoken by a distinct ethnic group mainly in Chad. Kanuri-Kanembu is part of the Saharan family, a well-defined subgroup within the Nilo-Saharan phylum.

Kanuri was intimately linked with the Kanem-Borno empire for almost 1000 years. Its role as the *lingua franca* of northern Nigeria was reduced in favor of Hausa during colonial times. Parallel to Hausa, a modified Arabic script known as Ajami was used in Kanuri for several centuries. The orthography based on Roman script and developed during the British colonial period was standardized in 1974. Today, Kanuri is used in mother-tongue education as well as on the radio and in television in Nigeria. It is also taught at university level, e.g., at the University of Maiduguri (Nigeria).

One of the earliest detailed analyses of Kanuri is Lukas (1937), who had already showed that Kanuri has two distinctive tone levels, low and high, as in *kànúrí* ('Kanuri person') or *kànúrí* ('Kanuri language'). These register tones may also be combined to build complex tones; the so-called 'mid' tone established by Lukas (1937: 3) represents a downstep high tone, i.e., a conditioned variant of the high tone. As further shown in Lukas' pioneering study, as well as in more recent analyses by Hutchison (1981), and by Cyffer and Geider (1997), Kanuri also has an intricate

system of consonant alternation, resulting in complex morphophonemic alternations.

One interesting typological property of Kanuri is the relatively small number of verbs. More than 95% of the predicative structures are formed by a combination of the light verb *n* ('say, think') and some complement (for example: *le+n* ('go')). This latter strategy is common in a range of Nilo-Saharan languages stretching from Kanuri in the Lake Chad region to Nara in Eritrea (see **Nilo-Saharan Languages**). Both complex predicates of this type and basic verbs are inflected for subject, tense-aspect-mood, as well as object (in the case of first and second person). Derivational argument modulation (more specifically, neutro-passive, causative, applicative marking) and pluractional marking is also expressed in the verbal complex.

Kanuri is head-final at the clausal level, involving verb-final structures, preverbal complementation, and the use of postpositions, but nominal modifiers follow the head noun. From a historical-comparative point of view, Kanuri grammar appears to be characteristic for the Saharan group within Nilo-Saharan. Interestingly, however, the lexical structure of Saharan languages appears to be less stable, as argued by Cyffer (2000).

Bibliography

- Cyffer N (2000). 'Linguistic properties of the Saharan languages.' In Zima P (ed.) *Areal and genetic factors in language classification and description: Africa south of the Sahara*. Munich: Lincom Europa. 30–59.
 Cyffer N & Geider Th (eds.) (1997). *Advances in Kanuri scholarship*. Cologne: Rüdiger Köppe.
 Cyffer N & Hutchison J P (1990). *Dictionary of the Kanuri language*. Dordrecht and Maiduguri: Foris Publications and University of Maiduguri.
 Hutchison J P (1981). *The Kanuri language: A reference grammar*. Madison: University of Wisconsin.
 Lukas J (1937). *A study of the Kanuri language*. London: Oxford University Press.

Kapampangan

H Kitano, Aichi University of Education, Aichi, Japan

© 2006 Elsevier Ltd. All rights reserved.

Kapampangan (Pampangan) is spoken mainly in Pampanga Province and in parts of Tarlac, Nueva Ecija, Bulacan, and Bataan provinces of Luzon, the Philippines. It is one of the largest languages in the Philippines. The number of its speakers is estimated as 1 897 378 (1990 census).

Phonology

Kapampangan has the consonants [p, b, t, d, k, g, ʔ, s, tʃ, ɕ, m, n, ŋ, l, r, w, y] and vowels [i, e, a, ø, u].

Overview of Kapampangan Grammar

Word Order

Like other Philippine languages, Kapampangan is a predicate-initial language. Clitic pronouns and clitic adverbs usually occupy the second position of a clause, and clitic pronouns are almost always obligatory, even when their coreferent noun phrases are present.

- (1) Masakit-ya ing lalaki
sick-ABS.3SG DET.ABS.SG man
'the man is sick'

Case Marking

Kapampangan has three cases: topic, genitive (or nontopic), and oblique. Since Kapampangan exhibits an ergative system in pronominal case marking, these cases may also be called absolutive, ergative, and oblique, respectively, which are the terms used in this article.

Negation

Sentential negation is expressed by the predicate-initial *e*.

- (2) E-ya masakit ing lalaki
NEG-ABS.3SG sick DET.ABS.SG man
'The man is not sick'

Existence and Possession

Both existence and possession are expressed by the existential particles *atin(g)* ('there is, be present, have') and *ala* ('there is not, be absent, do not have'). (LK=linker, realized either as *ng* [after vowels, *n*, and the glottal stop] or *a* elsewhere)

- (3) Ating metung a babai-ng malagu
EXIST one LK woman-LK beautiful
'there was a beautiful woman'

- (4) Atin-ya-ng kapatad a lalaki
EXIST-ABS.3SG-LK sibling LK man
'he/she has a brother (male sibling)'

Lexical Classes

Nouns Nouns may be monomorphemic or derived by affixation. A limited number of nouns may be pluralized with vowel lengthening (e.g., *lalá:ki* 'man' and *la:lá:ki* 'men,' *babá:i* 'woman' and *ba:bá:i* 'women').

Noun phrases are formed with other lexical classes, such as adjectives and verbs, by use of the linker.

- (5) ing bayu-mu-ng imalan
DET.ABS.SG new-ERG.2SG-LK dress
'your new dress'

- (6) ing balita-ku-ng dimdam
DET.ABS.SG news-ERG.1SG-LK heard
kang Mike
DET.OBL.SG Mike
'the news I heard from Mike'

Verbs Verbs derive for focus (actor, patient, directional, beneficiary, and instrumental; see section on focus constructions) and inflect for aspect (contingent [contemplated, future], perfective [completed, past], and imperfective [incompleted, progressive, present]). Verbs derive by way of affixation (prefixes, suffixes, and infixes), reduplication, vowel alternations, vowel lengthening, and combinations thereof. Some other affixes denote causative, aptative (abilitative, accidental, and coincidental actions), and distributive (states or actions distributed over space and time, and repetitive actions).

Adjectives Adjectives may be monomorphemic or derived, with affixes, such as *ma-* (plural *manga-*), which is the most common adjective-forming prefix. The comparative degree is marked by *mas* (e.g., *mas maragul* 'bigger'). Superlative adjectives are formed with the prefix *peka-* (*pekamaragul* 'biggest').

Determiners Determiners, also called articles, case markers, or noun markers, pronominally indicate the case (absolutive, ergative, or oblique) and number (singular or plural) of the nouns they qualify, and whether the nouns are common or personal (Table 1).

Pronouns Pronouns may be clitic or free. Some combinations of two clitic pronouns (ergative and

Table 1 Determiners in Kapampangan

| | <i>Absolutive</i> | <i>Ergative</i> | <i>Oblique</i> |
|-----------------------|-------------------|-----------------|----------------|
| Common nouns | | | |
| SG | ing, = ng | ning | king, keng |
| PL | deng/reng | reng | karing |
| Personal names | | | |
| SG | i | = ng | kang |
| PL | di/ri | ri | kari |

Table 2 Pronouns in Kapampangan

| | <i>ABS (clitic)</i> | <i>ABS (free)</i> | <i>ERG (clitic)</i> | <i>OBL (free)</i> |
|-----------|------------------------------|-------------------------------|-------------------------|---------------------|
| 1 SG | ku | yaku, aku | ku | kanaku, kaku |
| 2 SG | ka | ika | mu | keka |
| 3 SG | ya | iya | na | kaya |
| 1 DUAL | kata | ikata | ta | kekata |
| 1 PL INCL | katamu, tamu, kata, ta | ikatamu, itamu, ikata, ita | tamu, ta | kekatamu, kekata |
| 1 PL EXCL | kami, ke | ikami, ike | mi | kekami, keke |
| 2 PL | kayu, ko | ikayu, iko | yu | kekayu, keko |
| 3 PL | la | ila | da/ra | karela |

absolutive) may be fused into one word; e.g., *na* (ERG.1SG) + *ɣa* (ABS.3SG) becomes *ne* (Table 2).

Focus Constructions

In Kapampangan, as in other Philippine languages, the morphology of the verb indicates the semantic relationship between the predicate and the absolutive argument. In an actor-focus construction, the absolutive argument is semantically an actor, and the verbal predicate takes appropriate actor-focus affixes. Likewise, in a patient-focus construction, the absolutive argument is semantically a patient, and the verbal predicate takes patient-focus affixes. In the following examples of each focus construction, the boldfaced argument is the absolutive.

Actor focus:

- (7) Mamangan la-ng manuk
eating ABS.3 PL-LK chicken
'they are eating chicken'

Patient focus:

- (8) Kakanan de ing manuk
eating ERG.3PL + DET.ABS.SG chicken
 ABS.3SG
'they are eating **the** chicken'

Directional focus:

- (9) Dinan me-ng
give.to ERG.2SG + ABS.3SG-LK
pera ita-ng anak
money that.ABS.SG-LK child
'give some money to that kid'

Beneficiary focus:

- (10) Pangadi me
 pray.for ERG.2SG + ABS.3SG
 ‘pray for him/her/it’

Instrumental focus:

- (11) Penyulat ne
wrote.with ERG.3SG + ABS.3SG
ini-ng lapis
this.ABS.SG-LK pencil
'he/she wrote **with this pencil**'

Bibliography

- Forman M L (1971a). *Kapampangan grammar notes*. Honolulu: University of Hawaii Press.
- Forman M L (1971b). *Kapampangan dictionary*. Honolulu: University of Hawaii Press.
- Gonzalez A B (1981). *Pacific linguistics, series C, no. 48: Pampangan: towards a meaning-based description*. Canberra: Department of Linguistics, The Australian National University.
- Grimes B F (ed.) (2000). *Ethnologue* (14th edn.). Available at www.ethnologue.com.
- Kitano H (2001). 'Kapampangan.' In Garry J & Rubino C (eds.) *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*. New York/Dublin: H. W. Wilson Company. 371–375.
- Mirikitani L T (1971). *Speaking Kapampangan*. Honolulu: University of Hawaii Press.
- Mirikitani L T (1972). *Kapampangan syntax*. Honolulu: University of Hawaii Press.
- Mithun M (1994). 'The implications of ergativity for a Philippine voice system.' In Fox B & Hopper P J (eds.) *Voice: form and function*. Amsterdam: John Benjamins. 247–277.

Karen Languages

D B Solnit, University of Michigan, Ann Arbor, MI, USA

© 1994 Elsevier Ltd. All rights reserved.

Speakers of Karen languages make up one of the largest minority groups in both Burma and Thailand.

Classification

Karen belongs to the Tibeto-Burman side of the Sino-Tibetan family. In its subject-verb-object word order typology it stands apart from the subject-object-verb order of other Tibeto-Burman groups. This typological divergence is undoubtedly due to Karen's contact with SVO Mon-Khmer and Tai languages, and is not sufficient grounds for setting up Karen as member of Sino-Tibetan distinct from Tibeto-Burman, as proposed in Benedict (1972).

Location and Languages Included

Karen speakers are distributed along a north-south axis roughly coinciding with the Thailand-Burma border, reaching northwards into Shan State of Burma a bit beyond Taunggyi and southwards nearly to the Isthmus of Kra, with more scattered groups extending westwards into the Irrawaddy Delta and eastwards into Lampang and Chiang Rai Provinces.

A list of discrete Karen languages can only be approximate, but relatively well-defined languages include Sgaw and Pho in the southern portion and the east-west extensions of the area just described, and Pa-O (Taungthu) at the northern end. Less well-defined but still usefully referred to as unitary is Kayah (Red Karen, Karenni, Eastern Bwe), spoken in most of Kayah State and a few adjoining areas of Thailand. The center of Karen linguistic diversity is in western Kayah State and adjoining parts of Karen State, an area of complex dialect continua. Two languages of this area have been described, Palaychi by Jones (1961) and Western Bwe (Blimaw) by HENDERSON (1961). The remaining Karen languages include Padaung, located generally southwest of Pa-O and northwest of Kayah, and an indeterminate number of languages in the central area. All of these are known solely as a list of names (for details, including ethnographic notes, see Lehman 1967).

Number of Speakers

Anywhere from 3 to 4.5 million, of which perhaps 300 000 are in Thailand. Sgaw speakers are by far the most numerous in both Burma and Thailand.

Writing Systems

Writing systems akin to the Burmese script exist for Sgaw, Pho, and Pa-O, the former two having been developed by missionaries in the nineteenth century. Printed material includes Sgaw-English and Pho-English dictionaries. Other scripts, of less currency, exist for these and other Karen languages, with Thai, Roman, and eclectic sources.

Typological Characteristics

Karen languages share many features with other languages of mainland Southeast Asia, including contrastive tone, monosyllabicity, numeral classifiers, preference for aspect over tense, verb serialization, and lack of agreement, gender, and case marking.

One syllable per morpheme is the rule in Karen, although there are exceptions, notably those words including prefixes (see below). Karen tone systems typically have 3- and 4- way contrasts in syllables ending with vowel or sonorant, plus a 2-way contrast in syllables with final stops (if present as a distinct type). Tonal contrasts often include phonation as well as pitch features. The modern tone systems are the outcome of the conditioning effects of old initial consonant features acting on a proto-Karen system of 2 or 3 tones, very much as in Tai and Miao-Yao.

The basic Karen sentence type is verb-medial. Prepositions exist, although the repertoire is not large, with detailed spatial relations conveyed by noun expressions (e.g., 'at box's inside' for 'inside the box'). Nouns are modified by both preposed and postposed items; in general, nominal modifiers precede and verbal modifiers follow the head. The usual order for classifier constructions is noun-numeral-classifier. Word-formation is predominantly by compounding, although there are remnants of an earlier prefixation system, in the form of a collection of proclitic syllables with a more or less obscure morphemic identity.

Sample

The following represents the Eastern dialect of Kayah, spoken in Mae Hong Son Province of Thailand.

nè kè dá yè vè dī tənā tətā vè yò pō pā
you if give eat I rice one-day one-meal I guard IRREALIS
If you give me one meal a day to eat. I'll guard [them]

Bibliography

Benedict P K (1972). *Sino-Tibetan: A Conspectus*. New York: Cambridge University Press.

- Henderson E J A (1961). Tone and intonation in Western Bwe Karen. *Burma Research Society Fiftieth Anniversary Publication I*, 59–69.
- Jones R B (1961). *Karen Linguistic Studies*. Berkeley, CA: University of California Press.

- Lehman F K (1967). Burma: Kayah society as a function of the Shan–Burma–Karen context. In Steward J H (ed.) *Contemporary Change in Traditional Societies*, vol. 1. Urbana, IL: University of Illinois Press.

Kashmiri

S Munshi, University of Texas at Austin, Austin, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

Kashmiri, known to its speakers as *kə:ʃur/ko:ʃur* or *kə:ʃir zabān* ('Kashmiri language'), is spoken by around 4 million people in India (*Ethnologue*, n.d.), primarily in the Kashmir valley (*kə:ʃi:r*) and its surrounding regions in the state of Jammu and Kashmir (J & K). Sizeable population of Kashmiri speakers also live in other states of India. Approximately 0.1 million Kashmiri speakers have been reported to live in Pakistan (*Ethnologue*, n.d.). Kashmiri speakers are also found in various other countries. Major regional dialects of the language spoken in J & K state include: Standard Kashmiri (in and around Srinagar; used for educational and literary purposes), *Kashtwāri/Kishtwāri* (Kishtwār valley, southeast of Kashmir valley); *Poguli* (southern Banihal); *Sirāji* (several villages of district Doda); *Rāmbani* (Ramban and adjoining areas), and *Bunjwali*. *Poguli* and *Sirāji* are considerably influenced by *Dogri* (an Indo-Aryan language close to *Punjabi* (*Panjabi*)). Speakers of many of these dialects often argue for the separate existence of their native speech as full-fledged languages rather than dialects of Kashmiri. Information available on Kashmiri dialects is generally based on secondary sources.

There is a possibility of existence of more regional dialects in addition to the above-mentioned, but no detailed research on Kashmiri dialectology is available so far. Besides the major regional dialects, there are conspicuous differences between urban and rural Kashmiri on the one hand (in terms of accent, phonology, and lexicon; rural Kashmiri has preserved many archaic forms not used in urban speech) and between Muslim and Hindu Kashmiri on the other (mainly in terms of borrowings; while Hindus use more Sanskrit loans, Muslims commonly borrow from Persian (*Farsi*, Western) and Arabic for the corresponding words. The terms 'Persianized' and 'Sanskritized' Kashmiri are sometimes used to refer to these social dialects).

Historical Development

Kashmiri has been classified along with a number of languages grouped under the title 'Dardic.' Dardic languages are spoken in the extreme north of India and northwestern Pakistan, extending into Afghanistan. There has been considerable debate over the classification of Dardic languages as to whether they are a third branch of Indo-Iranian language family (other two being Indo-Aryan and Iranian), or (at least, some of them) are of pure Indo-Aryan origin. Dardic languages have preserved many archaic Indo-Iranian features otherwise lost in the modern Indo-Aryan languages. They have also developed certain features not found in other Indo-Iranian languages. Nevertheless, the term 'Dardic' constitutes a geographical convention rather than a linguistic expression. Like other Dardic languages, Kashmiri has similarities with both Indo-Aryan as well as Iranian. After continuing debates over a long period of time, many linguists have agreed upon an Indo-Aryan origin for Kashmiri. The term 'Dardic,' however, has gained much popularity, and is still used in view of the regional peculiarities shared by Kashmiri and other languages of the group.

Kashmiri belongs to the North-Western group of the Middle Indo-Aryan (MIA) languages/dialects, which includes several Dardic languages (e.g., *Shina*, *Khowar*, *Torwali*), *Punjabi*, *Sindhi*, and *Lahnda* (*Panjabi*, Western). One of the characteristic features with respect to the phonological system of the North-Western group is the retention of certain features lost elsewhere. In many modern Indo-Aryan languages, Old Indo-Aryan (OIA) sibilants – *ś* (palatal), *s* (dental) and *ṣ* (retroflex) – merged into dental *s*. Kashmiri retains two sibilants, the palatal *ś* and the dental *s*. Like other Dardic languages, Kashmiri also retains the consonantal component *r* in the derivatives of the OIA syllabic *ṛ*, which had a number of reflexes in MIA, viz., *a*, *i*, or *u*.

Kashmiri vocabulary can be broadly categorized into Kashmiri/Dardic, Sanskrit, Punjabi, Hindi/Urdu, Persian, and Arabic origins. Kashmiri occupies a special position in the Dardic group, being probably the

only Dardic language that has a written literature dating back to the early 13th century, a writing script of its own, and the largest number of speakers among the Dardic languages. An important part of Kashmiri is its rich tradition of oral and written literature. Some of the very famous genres of Kashmiri oral literature and folklore are *rov*, *vanivun*, *c^hakir*, *ladi: -śah*, and *luki-pə:t^hir*.

Writing Systems

Originally, Kashmiri was written in the Shāradā script, an ancient indigenous character of Kashmir. Shāradā is argued to be the predecessor of Devanāgri/Nāgri, which is built on the same system and corresponds with Shāradā letter-for-letter, although the letters have considerably changed in form. Shāradā is closely associated with Tākri alphabet used for writing Punjabi, but, with a complete range of symbols for different vowels characteristic of Kashmiri. Its use is highly restricted to a handful of Hindu priests in writing *za:tuk/Janam-patri* ('horoscope'). The most popularly employed and officially-recognized script in current use is a modification of Perso-Arabic (Nastālīq) script. Devanāgarī (again with modifications to cater to the specific requirements of the Kashmiri phonemic inventory) is also used and is popular among Hindus. Tākri (Kashtwāri and some dialects of the adjoining areas) and Roman scripts have also been employed but these have failed to gain recognition.

Phonological Characteristics

Both open and closed syllables are permitted in Kashmiri. Closed syllables, however, are preferred to open syllables. In rapid speech, in polysyllabic words with a sequence of adjacent CV syllables, speakers may drop medial vowels in favor of closed syllable structure. Final vowels are often deleted. Clusters comprising two consonants are quite common but only specific sequences can form a cluster in a particular position. Initial clusters are restricted to the type *Cr-* where the first consonant of the cluster is an obstruent. Final clusters are comprised of a homorganic nasal followed by an obstruent. The distribution of stress in Standard Kashmiri is influenced by a complex interplay of quantitative, positional, and rhythmic constraints. Primary stress appears on the word-initial syllable, which is always stressed. Stress occurs on every syllable containing a long vowel – CV:(C).

Kashmiri maintains the basic OIA pattern of five articulatory positions along with features of voice (e.g., *k* vs. *g*) and aspiration (e.g., *k* vs. *k^h*).

Its consonant system has survived with various alterations, losses, and additions. Some of the characteristic changes are those involving the loss of aspiration of voiced plosives, change of MIA palatal affricates to those of the corresponding dental affricates, word-final aspiration of voiceless stops, and fricative weakening/lenition (*s* > *h*) in certain environments. There are 27 consonantal phonemes in Kashmiri that have evolved from the Old/Middle Indo-Aryan phonological system. In addition, Kashmiri has also adopted a number of consonants from the Persian and Arabic phonemic inventory. The latter include the labio-dental voiceless fricative (*f*), voiceless and voiced velar fricatives (*x* and *ɣ*), and uvular and glottalic stops (*q* and *ʔ*). They are found only in Persian/Arabic borrowings and are used only by literate Kashmiris in Standard Kashmiri, especially in formal speech. In informal speech and among the illiterate population, these are replaced by *p^h*, *k^h*, *g*, *k*, and *a* respectively.

Like most of the New Indo-Aryan (NIA) languages, Kashmiri vowels are subject to various phonological operations that are not only regular but also extended over a larger domain than most of the other IA languages. A significant number of changes take place in accordance with the nature and position of the vowels in a particular linguistic domain (syllable/morpheme/word). Various such changes include: vowel harmony, *svarabhakti* (vowel epenthesis), consonantal assimilation, and final vowel deletion. Kashmiri has a 16-vowel system consisting of front vowels /*i*, *i:*, *e*, *e:*, *ɛ*/, central vowels /*ɨ*, *ɨ:*, *ə*, *ə:*, *a*, *a:*/, and back vowels /*u*, *u:*, *o*, *o:*, *ɔ*/ with three contrasts in height (high, mid, and low). Kashmiri phonemic inventory is distinct in the IA languages in having central vowels /*ɨ*, *ɨ:*, *ə*, *ə:*/ (absent in most NIA languages) and dental affricates.

Morphosyntax

Kashmiri, like other IA languages, is a postpositional language. However, its word order is unique among the IA languages. Unlike other IA languages, which are typically verb-final, Kashmiri is a V2 language. That is, the inflected verb occurs at the clause-second position. In sentences with a main verb and an inflected auxiliary verb, the main verb occurs sentence-finally. Thus, the basic word order is essentially SVO(V). This is true of the matrix as well as embedded clauses, and also of *yes-no* questions and questions where the *wh*-phrase is the syntactic 'subject.' In other question formations, the *wh*-phrase occupies the clause-second position, with the inflected verb occupying third position. In most environments, V2 is obligatory. However, in certain syntactic

environments, such as correlatives and conditional clauses, V2 is optional. The sentence initial position is occupied by syntactic 'subject' or any other constituent ('topic'). Except for the fixed position of the verb (and *wh*-phrase in case of questions), word order is fairly flexible. V2, characteristic of Germanic languages, is a well-developed syntactic phenomenon in Kashmiri, a non-Germanic language.

Complementation in Kashmiri is observed by optionally inserting complementizer *ki/zi* in front of the embedded clause. Relative clauses are formed by a relative-correlative construction with a pre-nominal relative pronoun in the modifying/relative clause and a correlative pronoun in the main clause following the relative clause. Relative-correlative construction is a typical areal syntactic feature.

Kashmiri is a split-ergative language. Case morphology is more or less typically Indo-Aryan. Based on thematic roles and verb class, the only argument of an intransitive clause receives either a zero/nominative case or a dative case. The most agent-like argument of a transitive clause in perfective aspect receives an ergative case and the other argument (if any) receives a nominative or dative case depending on its thematic role and the verb class. Subjects of a few intransitive verbs may appear in ergative case (e.g., verbs like *as-un* 'to laugh', *vad-un* 'to weep/cry'). Case ending behaves like a postposition so that the noun phrase appears in oblique form (a typical Indo-Aryan syntactic feature).

Kashmiri verb phrase is rich in agreement. Both subject and object agreement markers may appear on the verb. Subject agreement, however, is blocked in dative/ergative constructions where the syntactic subject is in dative/ergative case.

One of the characteristic features of Kashmiri among the Indo-Aryan languages is its three-way (instead of the typical two-way) distinction of the demonstrative pronoun, viz., (1) proximate *yi/yim* 'this/these', (2) visible *hu/hum* 'that/they (masculine)' and *ho/homa* 'that/they (feminine)', and (3) invisible/remote *su/tim* 'that/they (masculine)' and *so/tima* 'that/they (feminine)'.

Sociolinguistics of Kashmiri

Not many native speakers of Kashmiri can read the language, irrespective of their educational background. For a long time, Kashmiri has not been taught in schools. According to the three-language policy of India, languages generally taught in the schools of

J & K state are Urdu (a non-native language and the official language of the state), Hindi (official language of India), and English (second official language). There is a significant amount of prestige associated with Urdu, Hindi, and English in wider linguistic domains. These factors and increasing urbanization and globalization have played a significant role toward a gradual language loss with many Kashmiri speakers.

Sample (Srinagar/Standard Kashmiri)

| | | | | |
|-----|--|------------------------|-----------------------|---------|
| (1) | ga:ʃi-an | von | salim:as | (zi/ki) |
| | Gasha-ERG. | say.PAST.PERF. | Salim-DAT. | COMP |
| | agar | a:si-hund _k | mo:l _k | |
| | if | Asi-Gen.M.SG. | father | |
| | gar-i | o:s _k | tam-is | |
| | home-LOC. | be.PAST.M.SG. | 3SG.OBL.-DAT. | |
| | van-un | ma:j-i | samk ^h -un | |
| | tell-Infin. | mother.OBL.-DAT. | meet-INFIN. | |
| | 'Gasha _i told Salim _i (that) if Asi's father _k was home, he _i should tell him _k to meet (X's _{i/j/k/l}) mother' | | | |

Bibliography

- Bashir E (2003). 'Dardic.' In Cardona G & Jain D (eds.) *The Indo-Aryan languages*. London/New York: Routledge, Taylor & Francis Group. 818–894.
- Ethnologue (n.d.) 'Kashmiri'. *Languages of the World*, 15th edition.
- Grierson G A (1906). *The Pisaca languages of North-Western India*. London: The Royal Asiatic Society.
- Grierson G A (1919). *The linguistic survey of India VIII, Part II*. Calcutta: Superintendent Government Printing, India.
- Koul O N (2003). 'Kashmiri.' In Cardona G & Jain D (eds.) *The Indo-Aryan languages*. London/New York: Routledge, Taylor & Francis Group. 895–952.
- Koul O N & Hook P E (eds.) (1984). *Aspects of Kashmiri linguistics*. New Delhi: Bahri Publications Pvt. Ltd.
- Masica C P (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Morgenstierne G (1941). 'The Phonology of Kashmiri.' In *Acta Orientalia* 19(1). Reprinted in 1973 in *Irano-Dardica*. Wiesbaden: Dr. Ludwig Reichert Verlag. 79–99.
- Kachru B B (1969). 'Kashmiri and other Dardic Languages.' In Sebeok T A (ed.) *Current trends in linguistics 5: Linguistics in South Asia*. Paris/The Hague: Mouton. 284–306.

Relevant Website

http://www.ethnologue.com/show_language.asp?code=kas

Kayardild

N Evans, University of Melbourne, Melbourne, Australia

© 2006 Elsevier Ltd. All rights reserved.

Kayardild is spoken in Queensland, Australia, in the South Wellesley Islands, and belongs to the Tangkic family (non-Pama-Nyungan), which also includes Lardil and Yukulta. Additional and now extinct varieties are Yangkaal and Nguburindi, though the limited materials we have on these show Yangkaal to be a sister dialect of Kayardild and Nguburindi a sister dialect of Yukulta. The Tangkic languages have no close relatives, though they are related, at a distant level, to other Australian languages and share most grammatical similarities with languages along the Roper River well to the west (Evans, 1995).

Speakers of Kayardild were traditionally hunter-gatherers, with a strong emphasis on marine resources, building stone walls around the coasts to catch fish and hunting for turtle and dugong. The traditional population numbered between 120 and 150. Isolated from European contacts until the 1940s, the entire tribe was removed from their homeland in the early 1940s to the mission on Mornington Island, from which date rapid language loss set in: no one born after the move to Mornington grew up to be a fluent speaker, and today fewer than 10 speakers remain.

Apart from scanty early word lists, all materials on the Tangkic languages were recorded since the early 1960s. Practical orthographies were developed in this period. These use digraphs for a variety of phonemes, making use of *r* before a stop or nasal letter to denote retroflexion: thus *rd* for /ɾ/ and *rn* for /ɾŋ/; *h* after a stop or nasal letter to denote a laminoalveolar articulation (with the blade of the tongue between the teeth); thus *th* for /t̪/ and *nh* for /ɳ/. Other graphemes, standard in Australian orthographies, are *ng* for /ŋ/, *ny* for /ɲ/, *rr* for trilled or flapped /r/, *r* for /r/ and *j* for /ɟ/. Distinctive vowel length is shown by doubling the letter, e.g., *aa* for /aː/.

Phonologically, Kayardild is a typical Australian language, with paired stops and nasals at six points of articulation (bilabial, velar, laminoalveolar, lamino-palatal, apicoalveolar and apico-retroflex), a single stop series without voicing contrast, no fricatives, two rhotics (a glide and a tap/trill), and a simple vowel inventory: three vowels (*a*, *i*, *u*) plus length. Primary stress falls on the first syllable unless attracted onto a long vowel.

Kayardild is typical of Australian languages in employing a rich system of case suffixes, which allow for great freedom of word order. Beyond this, the case systems of Kayardild and the other Tangkic

languages are remarkable for several reasons. Firstly they exhibit ‘double case marking’ (see Dench and Evans, 1988; Plank, 1995), since one NP embedded in another inflects both for its own case (e.g., the possessive) and that of the head: cf. Kayardild *thabuju-karra* [brother-POSSESSIVE] ‘brother’s,’ *wangal-nguni* [boomerang-INSTRUMENTAL] ‘with the boomerang,’ *thabuju-karra-nguni wangal-nguni* ‘with brother’s boomerang.’ Secondly, Kayardild and Lardil add a further ‘modal case’ inflection, etymologically a case suffix, which marks tense/mood on most nonsubject NPs as a partly parallel system to the tense/mood inflection on the verb. Example (1) gives a Kayardild example using the ‘modal ablative’ (glossed M.ABL) to mark past tense on the object and instrument NPs in addition to the verbal ‘past’ inflection.

- (1) *dangka-a burldi-jarra yarbuth-ina*
man-NOM hit-PST bird-M.ABL
thabuju-karra-nguni-na wangal-nguni-na
brother-GEN-INST-M.ABL boomerang-INST-M.ABL
‘the man hit the bird with brother’s boomerang.’

Thirdly, Kayardild has a further ‘complementizing’ use of case suffixes, to indicate various types of interclausal relations such as being a clausal complement, as illustrated by the ‘complementizing’ use of the oblique in (2); note that it goes on all words of the subordinate clause, outside any other inflections. It is also used – on all words except the topicalized object – in strings of topic chains.

- (2) *ngada kurri-ja, dangka-ntha*
1sgNOM see-PST man-NOM
burldi-jarra-ntha yarbuth-ina
hit-PST-C.OBL bird-M.ABL-C.OBL
thabuju-karra-nguni-na wangal-nguni-na
brother-GEN- boomerang-
INST-M.ABL-C.OBL INST-M.ABL-C.OBL
‘I saw that the man had hit the bird with brother’s boomerang.’

Finally, Kayardild can add a layer of ‘associating case’ on all nonsubject NPs of clauses whose verb has been nominalized.

Much of the how this strange system evolved has now been reconstructed with the help of data from Yukulta, the most conservative Tangkic language and representative of the proto-Tangkic situation in having an ergative: absolutive case system; see Evans (1995) for a summary. Essentially, the main clause structures found in Kayardild and Lardil result from either reanalysis of alternative semitransitive structures in Yukulta, or from the generalization of Yukulta subordinate clause morphology, with case-marking interclausal relations, e.g., proprietive for purposive or ablative for prior time, smeared over

both the subordinate verb and its overt NPs. A series of catastrophic changes has thus led Kayardild and Lardil to have grammatical systems that are organized quite differently from that in Yukulta, despite the fact that almost all grammatical morphemes are cognate. These changes make the Tangkic languages a fascinating case of radical and intertwined diachronic developments linked to the abandonment of an ancestral system of ergative case marking.

Another strange feature of Kayardild is a further set of case inflections, semantically and structurally part of the set of 'normal' case inflections but with the peculiarity that they convert their hosts, morphologically, from nouns into verbs. Beneficiaries, for example, take the 'verbal dative' case *-maru-*, which then takes regular verbal inflections (3), but which is distributed across all words in the noun phrase like a case inflection. Etymologically this derives from a verb meaning 'put' but structurally it is now a part of the regular system of case suffixes.

- (3) ngada waa-jarra wangarr-ina
 1SgNOM sing-PST song-M.ABL
 ngijin-maru-tharra thabuju-maru-tharra.
 my-V.DAT-PST brother-V.DAT-PST
 'I sang a song for my brother.'

Like many other Australian languages, Kayardild has a complex set of derivatives from compass terms. To locate an entity one normally says things like 'the east uncle' or 'the groper coming from the east'; some examples of derivatives based on the root *ri-* 'east' are *riinda* 'coming from the east,' *rilungka* 'eastwards,' *riliida* 'heading ever eastwards,' *riyananganda* 'to the east of,' *ringurrnga* 'east across a geographical boundary,' *riyanyinda* 'at the eastern extremity of,' *rilumirdamirda* 'sea-grass territory to the east,' *rilur-ayaanda* 'from one's previous night's camp in the east,' *rilijulutha* 'move to the east' and *riinmali* 'hey you coming from the east.'

Bibliography

- Dench A & Evans N D (1988). 'Multiple case-marking in Australian languages.' *Australian Journal of Linguistics* 8, 1–47.
 Evans N D (1995). *A grammar of Kayardild, with historical-comparative notes on Tangkic*. Berlin: Mouton de Gruyter.
 Plank F (ed.) (1995). *Double case: agreement by Suffixaufnahme*. Oxford: Oxford University Press.

Kaytetye

H Koch, Australian National University, Canberra, Australia

© 2006 Elsevier Ltd. All rights reserved.

Kaytetye is the only member of the northern branch of the Arandic subgroup of the Pama-Nyungan language family, whose southern branch includes Arrernte (Aranda) and other languages (Hale, 1962; Koch, 2004). It is spoken in the southern part of the Northern Territory of Australia around Barrow Creek and Wauchope. Its speakers number around 200. Most contemporary people of Kaytetye descent

speak, in addition to or instead of Kaytetye, one of the neighboring languages – Alyawarr, Anmatyerre (both Arandic), or Warlpiri – and varieties of English. The current language of young people differs in a number of respects from that documented in the 1960s and 1970s by Hale and Koch.

Kaytetye phonology includes the typical Australian consonants, plus a set of pre-stopped nasals, rounded consonants, and a velar glide (the unrounded counterpart to *w*). An even more rare feature is a set of pre-palatalized apical consonants. The system of unrounded consonants and their orthographic representation is shown in **Table 1**. Rounded consonants

Table 1 Unrounded consonants (atypical segments in boldface)

| | Labial | Velar | Lamino-dental | Lamino-palatal | Apico-alveolar | Apico-postalveolar | Pre-palatalized apical |
|-------------------|-----------|------------|---------------|----------------|----------------|--------------------|------------------------|
| Stops | p | k | th | ty | t | rt | yt |
| Prestopped nasals | pm | kng | tnh | tny | tn | rtn | yt yn |
| Plain nasals | m | ng | nh | ny | n | rn | yn yl |
| Laterals | | | lh | ly | l | rl | |
| Tap/trill | | | | | rr | | |
| Approximants | | h | | y | | r | |

Table 2 Kin noun inflection

| | <i>'Elder brother'</i> | <i>'Mother'</i> |
|--------|------------------------|------------------|
| 1Sg | alkere-ye | arrwengke_ |
| 2Sg | ngk-alkere | ngk-arrwengke |
| 3Sg | kw-alkere | kw-arrwengke |
| Dyadic | alkere-nhenge | arrwengke-nhenge |

Table 3 Twelve words for 'we'

| | <i>Same moiety, same generation</i> | <i>Same moiety, opposite generation</i> | <i>Opposite moiety</i> |
|------------------|-------------------------------------|---|------------------------|
| Dual inclusive | ayleme | aylake | aylanthe |
| Dual exclusive | aylene | aylenake | aylenanthe |
| Plural inclusive | aynangke | aynake | aynanthe |
| Plural exclusive | aynenangke | aynenake | aynenanthe |

Table 4 Associated motion stems for intransitive and transitive verbs

| <i>Relative time of motion</i> | <i>Gloss</i> | <i>Angke 'talk'</i> | <i>Kwathe 'drink'</i> |
|--------------------------------|--|---------------------|-----------------------|
| Prior | VERB after SUBJ goes | angkeyene- | kwatheyene- |
| | VERB after SUBJ comes | angkeyethye- | kwatheyethye- |
| | VERB after SUBJ returns | angkeyalpe- | kwatheyalpe- |
| | VERB after non-SUBJ arrives | angkeyayte- | kwatheyayte- |
| Subsequent | VERB before SUBJ goes away | angkerrayte- | kwathelayte- |
| | VERB before SUBJ returns | angkerralpe- | kwathelalpe- |
| Concurrent | VERB while SUBJ comes | angkeyernalpe- | kwatheyernalpe- |
| | VERB while SUBJ goes along | angkerrape- | kwatherrapeyne- |
| | VERB continuously/repeatedly while SUBJ goes along | angkerrangkerrenye- | kwathelathelarre- |
| | VERB once while SUBJ is on the way | angkelpangke- | kwathelpathe- |
| Prior and subsequent | go and VERB and then return | angkenyayne- | kwathenyayne- |

are indicated by *w* after the consonant (cluster). The vowel system consists of just /a/, a high/mid-central vowel (spelled *e* in the orthography), and a marginal /i/. There is no rounded vowel phoneme, but /e/ has rounded allophones.

Atypical phonotactic features include common word-initial V(C) syllables and final CV syllables with no coda consonant and an obligatory vowel /e/; only word-internal syllables have the full structure CV(C). Word stress falls on the first CV(C) syllable. A word like *arrkwentyarte* 'three' illustrates these features (where the stressed syllable is underlined). The atypical phonology results from a series of sound changes that are described in Koch (1997, 2004).

Nouns inflect for number and case, in the typical Australian fashion, but with some complications (Koch, 1990). Kin nouns, as illustrated in Table 2, are inflected for the (singular) person of possessor, by means of prefixes (former dative pronouns) for second and third person, and a suffix or zero for first person. A suffix *-nhenge* marks the category 'dyadic',

designating both persons in a relationship, e.g., 'mother and child(ren)'.

Dual and plural personal pronouns, as seen in Table 3, mark a kinship-related category called 'section' (Koch, 1982), which distinguishes whether the participants belong to different patrimoiety (e.g., *I and my mother/spouse/sister's child*) or the same moiety; and if the latter, to the same or opposite set of alternate generation levels (e.g., *I and my siblings/grandparents/grandchildren* vs. *I and my father/brother's children*).

Verbs may inflect for the category of 'associated motion' (Koch, 1984), indicating distinctions in the direction and relative timing of movement by the subject (usually), using markers that partially differ according to the transitivity of the verb stem (Table 4) and derive in part from former verbs of motion.

In semantics, the expression of feelings and emotions is characterized by the use of the reflexive construction of 'hear' and the mention of body parts, especially *aleme* 'stomach' (Turpin, 2002).

Rewenhe aleme eyterrye-le elpathe-nke errpatye
 3RDSING.REFL stomach person-ERG hear-PRES bad
 'That person feels bad'

Available language resources include a non-technical learner's guide (Turpin, 2000) and picture dictionary (Turpin and Ross, 2004), each of which includes an audio component, and a text collection (Thompson, 2003).

Bibliography

- Hale K (1962). 'Internal relationships in Arandic of Central Australia.' In Capell A (ed.) *Some linguistic types in Australia* (Oceania Linguistic Monographs 7). Sydney: University of Sydney. 171–183.
- Koch H (1982). 'Kinship categories in Kaytej pronouns.' In Heath J, Merlan F & Rumsey A (eds.) *Languages of kinship in Aboriginal Australia* (Oceania Linguistic Monographs 24). Sydney: University of Sydney. 64–71.
- Koch H (1984). 'The category of 'associated motion' in Kaytej.' *Language in Central Australia* 1, 23–34.
- Koch H (1990). 'Do Australian languages really have morphemes? Issues in Kaytej morphology.' In Austin P, Dixon R M W & Dutton T (eds.) *Language and history: essays in honour of Luise A. Hercus* (Pacific Linguistics C-116). Canberra: Australian National University. 193–208.
- Koch H (1997). 'Pama-Nyungan reflexes in the Arandic languages.' In Tryon D & Walsh M (eds.) *Boundary rider: essays in honour of Geoffrey O'Grady* (Pacific Linguistics C-136). Canberra: Australian National University. 271–302.
- Koch H (2004). 'The Arandic subgroup of Australian languages.' In Bowern C & Koch H (eds.) *Australian languages: classification and the comparative method*. Amsterdam/Philadelphia: John Benjamins. 127–150.
- Thompson T K (2003). *Growing up Kaytetye: stories*. Turpin M (comp.). Alice Springs: IAD Press.
- Turpin M (2000). *A learner's guide to Kaytetye*. Alice Springs: IAD Press.
- Turpin M (2002). 'Body part terms in Kaytetye feeling expressions.' In Enfield N J & Wierzbicka A (eds.) *The body in description of emotion*. Amsterdam/Philadelphia: John Benjamins. 271–305.
- Turpin M & Ross A (2004). *Kaytetye picture dictionary*. Alice Springs: IAD Press.

Kazakh

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Kazakh (*qazaq tili*, *qazaqša*) belongs to the north-western or Kipchak branch of the Turkic language family, more specifically to its southern or Aralo-Caspian group. Until the early 20th century, it was called Kazak-Kirghiz, whereas Kirghiz was referred to as Kara-Kirghiz. Kazakh is primarily spoken in the Republic of Kazakhstan (*Qazaqstan Respublikası*), a vast country situated at the center of the West Eurasian steppe zone. It borders on Turkmenistan, Uzbekistan, Kyrgyzstan, and China in the south, and on the Russian Federation in the north and west. Kazakh is also spoken by minorities in Xinjiang (China), Uzbekistan, Mongolia, Turkmenistan, Kyrgyzstan, the Russian Federation, Tajikistan, Afghanistan, etc. The number of speakers is at least 10 million. There are more than seven million in Kazakhstan, more than one million in Xinjiang, and almost one million in Uzbekistan.

Kazakh is, along with Russian, the official language of the Republic of Kazakhstan. Kazakh–Russian

bilingualism is widespread. Though Kazakhs constitute half of the population of the republic, many have a low proficiency in their mother tongue. Russians make up 37% of the population. The declaration of Kazakh as the state language in 1989 was met by protests from the non-Kazakh population. In 1995, Russian was proclaimed the language of interethnic communication. Russian has a dominant status in public life as the main language of instruction, science, business, and communication in professional domains.

Karakalpak (*qaraqalpaqša*), an independent language in the political sense, is a slightly Uzbekicized variety of Kazakh. It is spoken by c. 450 000 persons, mainly in the Autonomous Republic of Karakalpakstan (*Qaraqalpaqstan Respublikası*) in Uzbekistan, on the lower course and in the delta of the Amudarya River. Small groups of speakers live in other regions, e.g., in the Khorezm and Fergana regions of Uzbekistan and in the Dashkhowuz region of Turkmenistan.

Origin and History

Kazakh goes back to the Kipchak varieties of Uzbek tribes, who founded a huge steppe empire in the second part of the 15th century. Separatist Kazakh tribes split off from the Uzbeks during their

migrations, and moved into the northern steppe regions. The ancestors of the Karakalpaks belonged to one of the most important confederations under the Golden Horde. At the beginning of the 17th century, the Kazakhs occupied Tashkent, which remained their capital until 1723. In the 18th century, Kazakh and western Mongol tribes fought for supremacy in the steppes between the Altay mountains and the Caspian Sea. The Kazakh empire disintegrated into three so-called hordes, of which the Great Horde submitted to Russia in 1717. The Kazakh territory was incorporated into Russia in the mid-19th century. A Kazakh constituent republic of the Soviet Union was established in 1920. Kazakhstan declared its sovereignty in 1990 and its full independence in 1991.

Related Languages and Language Contacts

Kazakh is closely related to Karakalpak, Kipchak Uzbek and Noghay (Nogai). Kipchak Uzbek, a vanishing variety formerly spoken mainly in the north and northwest of Uzbekistan, goes back to the original language of the Uzbek nomads. Kazakh has had old close contacts with Mongolic languages, and the recent influence on it of Russian has been strong. The southern part of Kazakhstan constitutes an intensive contact zone with Uzbek (Northern Uzbek). The Kazakh varieties spoken in China have been relatively little influenced by Chinese.

The Written Language

Kazakh did not possess a written variety in the pre-Russian period. Official documents were mostly written in Chaghatay (Chagatai) or Tatar. At the end of the 19th century a written language emerged. It was based on the dialect of the northwestern regions, where the Russian and Tatar influence was strong.

The Arabic script was used up to 1929. It was replaced by a Roman-based alphabet, which was abandoned in 1940 in favor of a Cyrillic alphabet. There are currently plans to adopt a Roman-based script again. In China, Kazakh is written in Arabic script, after an unsuccessful experiment with a Roman-based (pinyin) alphabet in the 1970s. Written Kazakh of China is still oriented towards the standard language used in Kazakhstan.

Karakalpak was established in 1925 as a language in the political sense and as a written language. Its orthography differs considerably from that of Kazakh. After the Arabic and Roman scripts had been employed for a few years, the Cyrillic alphabet was introduced in 1940. The transition to a new Roman-based script has begun.

Distinctive Features

Kazakh exhibits most linguistic features typical of the Turkic family (see **Turkic Languages**). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation, e.g., *A* = *a/e*. Hyphens are used here to indicate morpheme boundaries.

Phonology

Kazakh has a front vowel *æ*, which is lower than the mid-high *e* and restricted to the first syllable. It has emerged through fronting of *a*, e.g., *bæri* 'all.' High vowels are often reduced, relatively short, but not lowered as in Tatar and Bashkir. Initial *e-*, *o-* and *ö-* are often pronounced with a prothetic glide, e.g., *ʔeki* 'two,' *ʔon* 'ten.' Initial *t-* is mostly preserved, e.g., *til* 'language' (cf. Turkish *dil*). As in most Kipchak languages, final *-y* is labialized in monosyllabic stems, e.g., *taw* 'mountain' ← *ta:y*. The affricate *č* has developed into the fricative *š*, e.g., *üč* 'three' ← *üšč*, whereas *š* has developed into *s*, e.g., *tas* 'stone' ← *taš*. Word-initial *ž-* corresponds to *y-* or *ǰ-* in other Turkic languages, e.g., *žol* 'way' (Turkish *yol*). However, *ǰ-* is found in older Kazakh texts.

According to front vs. back sound harmony, a suffix vowel is back if the preceding syllable has a back vowel, and front if the preceding syllable has a front vowel. This type of sound harmony also affects consonants, e.g., *qar-ya* [snow-DAT] 'into the snow,' *köl-ge* [lake-DAT] 'into the lake.'

According to rounded vs. unrounded harmony (labial harmony), a suffix vowel is rounded if the preceding syllable contains a rounded vowel, and unrounded if the preceding syllable contains an unrounded vowel. Spoken Kazakh displays this kind of harmony not only in high suffix vowels, but also in low vowels, e.g., *üy-dö* [house-LOC] 'in the house,' *tüs-kön* [fall-PART] 'fallen,' *öl-gön* [die-PART] 'dead.' However, *o* is not admitted, e.g., *qol-da* [hand-LOC] 'in the hand' instead of **qol-do*. The rounding effect decreases with the distance from the first syllable, e.g., *üy-dö* [house-LOC] 'in the house,' *üy-ümüz-de* [house-POSS.1PL-LOC] 'in our house.' The rounded vs. unrounded harmony is not reflected in the orthography, which does not represent the suffix vowels *ö*, *ü*, and *u*.

Kazakh exhibits numerous consonant changes, mostly assimilations, in clusters containing dentals, liquids and nasals. Suffix-initial *d-*, *l-*, *n-*, and *m-* occur after stem-final vowels and often after nasals, sonorants, and glides, but they are otherwise assimilated. Suffix-initial *d-*, *l-*, and *n-* are assimilated

to *t-* after voiceless stem-final consonants, e.g., *at-tar* [horse-PL] ‘horses’ ← *at-lar*. Suffix-initial *d-* is assimilated to *n-* after stem-final nasals, e.g., *adam-nan* [man-ABL] ‘from the man.’ Suffix-initial *l-* and *n-* are changed to *d-* after most consonants, e.g., *köz-der* [eye-PL] ‘eyes,’ *qız-dı⁹* [girl-GEN] ‘of the girl.’ Suffix-initial *m-* is changed to *b-* after stem-final *-z* and *-ž*, e.g., *žaz-ba-* [write-NEG] ‘not to write.’

Loanwords are mostly pronounced according to indigenous phonotactic rules. Consonant clusters are dissolved through consonant deletion, or insertion of high prothetic or epenthetic vowels, e.g., *xaliq* ‘people’ (Arabic *xalq*), *iras* ‘true’ (Persian *rast*), *pıratsent* ‘percent’ (Russian *procent*), *telgirep* ‘telegraph’ (Russian *telegraf*), *kerewet* ‘bed’ (Russian *krovat*). The consonants *f* and *x* in loanwords are replaced by *p* and *q*, e.g., *aptobus* ‘bus,’ *qat* ‘letter.’ The phonological adaptation was largely reflected in the old Arabic and Roman orthographies, but since the introduction of the Cyrillic script, Russian loanwords are written in their original form.

Grammar

The comparative suffix is *-rAK* (after consonant-final stems *-(l)rAK*), e.g., *ülken-irek* [big-COMP] ‘bigger.’ There is an instrumental case in *-Men*. The reflexive pronoun *öz* is used attributively or as a noun with possessive suffixes, e.g., *öz üy-im* [self house-POSS.1SG] ‘my own house,’ *öz-im bar-dım* [self-POSS.1SG go-PAST-1SG] ‘I went myself.’

The old flexion of pronouns is preserved, and not replaced by flexion of the nouns as in some other Turkic languages. Pronouns ending in *-l*, e.g., *bul* ‘this,’ replace *-l* with *-n* in most oblique cases. The initial *b-* in *bul* is changed to *m-* in the genitive, accusative, and locative. The demonstrative pronouns *bul*, *osı*, and *mına* are used for referents within the range of view, while *ol*, *sol*, and *ana* are used for referents out of sight. Some demonstratives exhibit emphatic forms, e.g., *mına-w*.

The second-person plural marker *sIn-dAr* combines the second-person singular marker with the plural suffix. The marker *-sIz* expresses politeness when used to one addressee, whereas *-sIz-dAr* is the corresponding polite plural marker.

The suffixes *-LA-p* and *-LA-GAn*, whose first element *-LA* derives verbs stems from nouns, form approximative and multiplicative numerals, e.g., *žüz-de-gen* ‘hundreds.’

Kazakh has a present tense in *-A* plus person markers, e.g., *kel-e-di* [come-PRES-3.SG] ‘comes.’ It has emerged from a construction containing a converb ending in *-A + tur(ur)* ‘stands’. The old present tense form in *-(A)r* is mostly used with modal meanings,

e.g., *kör-er-miz* [see-AOR-1.PL] ‘we will see.’ A more focal present tense, i.e. with a narrower focus on the ongoing event, is formed with the converb suffix *-(l)p* and one of the verbs *otır* ‘sits,’ *žür* ‘goes,’ *žatır* ‘lies,’ *tur* ‘stands,’ e.g., *žaz-ıp žatır* [write-CONV lies] ‘is writing.’ The suffix *-AtIn* (*-ytIn* after vowel-final stems) forms a habitual past. There is an intentional in *-MAK*, e.g., *kel-mek-piz* [come-INTENT-1PL] ‘we want to come,’ and a necessitative in *-MAK-šI*, e.g., *sat-paqši-mün* [sell-NEG-1SG] ‘I must sell.’ Kazakh displays complex verbal compositions expressing actionality, aspect-tense, and evidentiality. A number of auxiliary verbs are used in postverb constructions based on the converbs in *-A* and *-(l)p* and mostly specifying the manner of action, e.g., *žan-ıp ket-* [burn-CONV go] ‘to burn down.’ Evidential forms are the indirective past in *-(l)p*, with *-DI* in the third person, e.g., *kel-ip-ti* [come-CONV-3SG] ‘apparently came.’ The evidential (indirective) copula particle *eken* combines with various items, e.g., *kel-e-di eken* [come-PRES-3SG INDIR. COP] ‘apparently comes,’ *kel-gen eken* [come-PART EV] ‘has apparently arrived.’ Unlike more Iranicized languages such as Uzbek and Turkmen, Kazakh has a weakly developed system of conjunctions.

Lexicon

The basic vocabulary related to the traditional way of life is of Kipchak Turkic origin. There is also a rich modern Kazakh vocabulary with numerous neologisms, e.g., *xaliqaraliq* ‘international.’ Many words are of Persian (Farsi) and Arabic origin, introduced via Tatar and Chaghatay, e.g., *bazar* ‘market,’ *apta* ‘week,’ *nan* ‘bread,’ *aqıl* ‘intellect,’ *ğilim* ‘science,’ *mayına* ‘meaning,’ *waqıt* ‘time.’ Due to close contacts with other nomadic groups, Kazakh also exhibits many words of Mongolic origin, e.g., *olža* ‘booty.’ Kazakh of China has a fairly strong early layer of Mongolic loanwords, but relatively few loans from Chinese, at least in the written language. Russian words have been borrowed from the second half of the 19th century on. In the Soviet period, calques on Russian models and neologisms were generally preferred to loanwords. There is now a certain tendency in the written language to reduce the number of Russian loans in favor of words of native or Arabic–Persian origin.

Dialects

In spite of its huge extent, the Kazakh-speaking area exhibits little dialectal variation because of a high degree of mobility of the speakers of Kazakh throughout their history. The standard language is based on the northwestern dialect. The southern

and western dialects differ from it in some respects. For example, initial *y-* is often found, e.g., *yaq* 'side' instead of *žaq*. Common Turkic *č* is often found instead of standard Kazakh *š*. Changes of suffix-initial consonants are less common. They are, on the other hand, stronger in the easternmost dialects spoken in China and Mongolia, e.g., *bala-dar* [child-PL] 'children' instead of *bala-lar* [child-PL]. The rural Kipchak Uzbek dialects in the north and northwest of Uzbekistan have been largely de-Kipchakicized and are now practically extinct due to the abandonment of the nomadic lifestyle.

The Karakalpak language displays most of the phonetic and morphophonemic characteristics of Kazakh. Word-initial *ǰ-* is, however, found instead of *ž-*, e.g., *ǰol* 'road.' Suffix-initial *l-* is not assimilated, e.g., *tas-lar* [stone-PL] 'stones' instead of *tas-tar*, *qız-lar* [girl-PL] 'girls' instead of *qız-dar*. There are a few morphological differences, stemming from the influence of Oghuz and Uzbek. The future suffix *-AžAK* is a loan from Oghuz. The first- and second-person plural personal markers *-MAN* and *-SAN* instead of Kazakh *-MIn* and *-sIN* are similar to Uzbek *-man* and *-san*. The vocabulary differs to a certain degree from that of Kazakh and contains more words of Arabic and Persian origin. Karakalpak has two main dialects, one northeastern and one southwestern. Certain nominally Karakalpak and Kazakh dialects

in the region of Khorezm belong to the Southwestern or Oghuz branch of Turkic.

Bibliography

- Balakaev M B *et al.* (eds.) (1962). *Sorremennij kazaxskij jazyk*. Alma-Ata: Izdatel'stvo Akademii Nauk Kazaskoj SSR.
- Johanson L (2001). 'Kazakh.' In Garry J & Rubino C (eds.) *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*. New York: H. W. Wilson. 376–378.
- Kirchner M (1992). *Phonologie des Kasachischen: Untersuchungen anhand von Sprachaufnahmen aus der kasachischen Exilgruppe in Istanbul*. Wiesbaden: Otto Harrassowitz.
- Kirchner M (1998). 'Kazakh and Karakalpak.' In Johanson L & Csató É Á (eds.) *The Turkic languages*. London/ New York: Routledge. 318–332.
- Krueger J R (1980). *Introduction to Kazakh*. Bloomington: Indiana University Press.
- Menges K H (1959). 'Die aralo-kaspische Gruppe.' In Deny J *et al.* (eds.) *Philologiae turcicae fundamenta 1*. Wiesbaden: Steiner. 434–488.
- Mirzabekova Q *et al.* (eds.) (1992). *Qazaqşa-nemisşe sözdik/ Kasachisch-deutsches Wörterbuch*. Almaty: Rawan.
- Shnitnikov B N (1966). *Kazakh-English dictionary*. London/The Hague/Paris: Mouton.
- Wurm S (1951). 'The Karakalpak language.' *Anthropos* 46, 487–610.

Keres

M J Mixco, University of Utah, Salt Lake City, UT, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Of the 21 Pueblo tribes in the American Southwest, one speaks Zuni, a language isolate; another speaks Hopi (Uto-Aztecan family) in several Arizona villages. In New Mexico, Tanoan pueblos (speaking Tiwa, Tewa, or Towa, related to Kiowa) are interspersed with those speaking Keres (or Keresan, *Quirix* or *Quires* in 16th-century Spanish; Davis, 1959).

Historically, Keres borrowed lexically from other pueblo languages via bilingual speakers, now a defunct process. Keresan has two major dialect divisions: Eastern (Rio Grande Valley) and Western (Colorado River drainage). Mutual intelligibility among dialects increases with proximity (speaker populations follow in parentheses as cited in Mithun, 1999, from Valiquette, 1995), Western: Acoma

(1930)/Laguna (2060); Eastern: Zia (504)/Santa Ana (384)-San Felipe (1985)/Santo Domingo (2965)-Cochiti (525).

Beginning in 1598, Spanish influences were strong only in the area of material culture (livestock, new crops, trade goods, and technologies). The impact was slighter in nonmaterial domains (e.g., religion) because of indigenous resistance resulting in the Pueblo Revolt of 1680–1692.

Surrounded by denser Spanish populations, the Eastern dialects received the most loanwords via Spanish bilingualism that persisted into the mid-20th century. English displaced Spanish as the regional lingua franca but provided fewer loans. Keres linguistic literature is scant, despite intensive Pueblo ethnographic work.

Structural Overview

Keres consonantal series are plain (b, d, d^y, g), aspirated (p, t, tʃ, k), and glottalized (p', t', tʃ', k') stops

and affricates; sonorants (m, n, r, w, y, m', n', r', w', y'); and fricatives (s, ʃ, f, s', ʃ', f'). There are two laryngeals (ʔ, h); five long and short vowels, voiced and voiceless contextually (i, e, a, ɪ, u); and as many as four pitch accents (high ˊ, falling ˆ, breathy ˋ, and glottal ˑ), depending on the dialect. Morphologically, Keres is polysynthetic, exhibiting distinct sets of prefixes on active and stative verbs to express nominal arguments without independent pronouns. Pronominal prefixes, with the subject and object fused into one morpheme, undergo especially complex phonological alternations and simultaneously encode five subject and object persons, including an indefinite and a fourth or obviative person (when the subject is hierarchically lower than the object). Also simultaneous is modality (negative, dubitative, hortative, negative hortative, future hortative, and indicative). Singular, dual, and plural numbers occur. Voice requires its own set of pronominal prefixes followed by the reflexive-reciprocal prefix /-a-/ or the passive /-aʔa- or -aʔ-/ . Benefactives are expressed discontinuously by prefixes and suffixes. Aspect (continuous, fulfilled, and state) is suffixal, as is adverbial subordination. Some intransitive pronominal agreement examples are [ʃ-uʔpɪ] 'I or you ate,' [k-uʔpa] 'he ate' versus the transitives [sʔ-ʔaku] 'I bit you', [tʃʔ-ʔaku] 'you bit me,' [ʃ-ékʊ] 'I bit (him),' [sʔ-ʔaku] 'he bit (him),' [sg-àku] 'someone bit him' (Mithun, 1999: 438–440). The following sentences are from Davis (1964) as quoted in Mithun (1999):

ʔe ʃu ʔe yúsi nʔ ʔeu dya:mi ʔeu ʃu ʔe yúsi
 nʔ tʔ-a'gúyan-e
 from there down EMPH eagle from there down 3.DUB-
 PASS-send-PL
 'then the eagle was sent down from above'
 ʃu ʔe hauʔ dí-uw'ác'i
 PRT PRT toward 3.DUB-approach
 '(the eagle) approached'

Language, Culture, and Society

The Keres language reflects the traditional inter-related cultural concerns of the Pueblo peoples, principally religious ceremonialism, agriculture, and theocratic governance. This has given rise to elaborate esoteric terminologies distinguishing the realms of the sacred from the profane. The latter categories

may vary from pueblo to pueblo. There is baby talk as well as a difference between male and female speech in frequently occurring words (Kroskrity, 1983; Sims and Valiquette, 1990).

Recently, there have been vigorous efforts in some Keresan pueblos to counteract language shift (Pecos and Blum-Martinez, 2001; Sims, 2001).

Bibliography

- Campbell L (1997). *American Indian languages: the historical linguistics of Native America*. Chicago: University of Chicago Press.
- Davis I (1959). 'Linguistic clues to northern Rio Grande prehistory.' *El Palacio* 66, 73–84.
- Davis I (1963). 'Bibliography of Keresan linguistic sources.' *IJAL* 29, 289–293.
- Davis I E (1964). 'The language of Santa Ana.' *Bureau of American Ethnology Bulletin* 191, 53–190.
- Davis I E (1979). 'The Kiowa-Tanoan, Keresan, and Zuni languages.' In Campbell L & Mithun M (eds.) *The languages of Native America: historical and comparative assessment*. Chicago: University of Chicago Press.
- Kroskrity P V (1983). 'On male and female speech in the Pueblo Southwest.' *IJAL* 49, 88–91.
- Miller W R (1965). *Acoma grammar and texts* (vol. 40). University of California Publications in Linguistics, Berkeley: University of California Press.
- Miller W R & Davis I E (1963). 'Proto-Keresan phonology.' *IJAL* 29, 310–330.
- Mithun M (1999). *The languages of native North America*. Cambridge Language Surveys. Cambridge, UK: Cambridge University Press.
- Pecos R & Blum-Martinez R (2001). 'The key to cultural survival: language planning and revitalization in the Pueblo de Cochiti.' In Hinton L & Hale K (eds.) *The green book of language revitalization in practice*. New York: Academic Press.
- Sims C P (2001). 'Native language planning: a pilot process in the Acoma Pueblo community.' In Hinton L & Hale K (eds.) *The green book of language revitalization in practice*. New York: Academic Press.
- Sims C P & Valiquette H (1990). 'More on male and female speech in (Acoma and Laguna) Keresan.' *IJAL* 56, 162–166.
- White L A (1944). 'A ceremonial vocabulary among the Pueblos.' *IJAL* 10, 161–167.
- Valiquette H P (1991). 'A study for a lexicon of Laguna Keresan.' *Dissertation Abstracts International, A: The Humanities and Social Sciences* 52(2), 527-A.

Ket

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

Ket is also known as Yenisei Ostyak and Imbat[skij] Ket. There are probably fewer than 200 mother tongue speakers out of an ethnic population of 1100–1200 (Krivonogov, 1998). Ket is spoken in north central Siberia along the Yenisei river and its tributaries (e.g., Yelogui), in northern Krasnoyarski Kray, spoken mainly in the villages of Sulomai, Kellog, Surgutikha, and Maiduka. Ket is the sole surviving member of the Yeniseic language family. Related languages include Arin, Assan, and Pumpokol, which became extinct by mid-18th century; Kott, which lasted until the mid-19th century, originally spoken to the north, west, and south of Krasnoyarsk. The closely related Yugh (also known as Sym-Ket) became extinct only in 1980s, and was last spoken in Vorogovo village. Yeniseic language speakers shifted to Russian or various Turkic varieties, in particular Chulym Turkic, Xakas, Shor, and Northern Altai varieties. Yeniseic speakers previously occupied the territory south along the Yenisei to the mouth of the Dupches River. Yeniseic-speaking peoples once occupied a large area in western and central Siberia, based on the widespread use of Yeniseic hydronyms across the area. Presumably, the attested Yeniseic peoples were encroached upon and marginalized areally by the advance at various periods of Samoyedic- and Ob-Ugric-speaking peoples from the west, and Turkic and Tungusic from the south and east, until they occupied their attested position along the Yenisei. The Xiong Nu of the Chinese chronicles may have spoken a Yeniseic language.

There are three dialects of Ket: Northern, Central, Southern. Mainly Southern Ket survives in such villages as Kellog and Sulomaj. Ket is severely endangered but enjoys high status in the villages where Ket people dominate. Only a handful of young people are learning the language as a first language, but Ket instruction in primary schools has begun in certain villages (e.g., Kellog), based on the written form of Ket developed by native Ket scholars and Professor Heinrich Werner.

Ket is unique among central Siberian languages for its unusual system of tone. Tones appear as a prosodic feature of the two leftmost syllables (if present) in Ket words (Vajda, 1999).

- (1) Southern Ket
- | | |
|----------------------------|----------------|
| ¹ <i>sʲulʲ</i> | ‘blood’ |
| ² <i>sʲuʲlʲ</i> | ‘white salmon’ |
| ³ <i>sʲu:lʲ</i> | ‘sled’ |

⁴*sʲulʲ* ‘cradle hook’
(Vajda, 1999: 5)

Tone differentiates both lexical and grammatical forms in Ket.

- (2) Southern Ket
- | | | | |
|----------------|----------------|----------------|-----------------------------------|
| <i>ásəl</i> | ‘ski’ | <i>àsél</i> | ‘type of large covered houseboat’ |
| <i>bʰántán</i> | ‘mallard duck’ | <i>bʰántàn</i> | ‘mallard ducks’ |
- (Vajda, 1999: 13)

The Ket dialects differ tonally primarily in the realization of the fourth tone, which may be falling with a short vowel and no pharyngealization (Southern Ket), or pharyngealized with a long vowel (Central and Northern Ket).

- (3) Southern Ket Central Ket Northern Ket gloss
- | | | | |
|---------------------------|-----------------------------|-----------------------------|------------|
| ⁴ <i>sʲɛlʲ</i> | ⁴ <i>sʲɛ:lʲi</i> | ⁴ <i>sʲɛ:lʲi</i> | ‘reindeer’ |
| ⁴ <i>asʲ</i> | ⁴ <i>a:sʲe</i> | ⁴ <i>a:sʲe</i> | ‘feather’ |
| ⁴ <i>ɪrʲ</i> | ⁴ <i>i:də</i> | ⁴ <i>i:rʲe</i> | ‘spring’ |
- (Vajda, 1999: 7)

Although Ket is unique among the languages of Siberia for its tone/register system (certain Ket tones, e.g., tone 2, are associated with pharyngeal tension), Ket is typical of north-central Siberian languages for its elaborate case system (Anderson, 2003). Nouns may appear in one of ten or eleven case forms. Ket is unusual for Siberian languages in distinguishing three genders or noun classes, roughly masculine, feminine, and other/neuter. One set of the case forms encodes these distinctions. Others do not and appear to be more recently fused postpositional constructions.

- (4) Ket
- | | |
|------------------|---------------------|
| <i>bʰib-dəŋa</i> | <i>bʰun-dʲəŋa</i> |
| son-DATIVE. | daughter-DATIVE. |
| MASCULINE | FEMININE |
| ‘to (his) son’ | ‘to (his) daughter’ |
- (Werner, 1997: 105)

In addition, Ket makes extensive use of case forms on verbs as well within a highly diversified system of case-marked clausal subordination that dominates complex sentence structure in the languages of the region.

- (5) *-dʲəŋalʲ* Ablative → ‘after,’ ‘since’
- | | | |
|------------|----------------|---------------------------|
| Ket | | |
| <i>bu</i> | <i>ətn-asʲ</i> | <i>du-y-a-raq-dʲəŋalʲ</i> |
| he | we-INS/COM | I-SEP-PRES-live-ABL |
| <i>dəŋ</i> | <i>sikŋ</i> | <i>u-γəŋ</i> |
| three | year.PL | III-go |
- ‘three years have passed since he’s been living with us’
(Werner, 1997: 353)

Among the most noteworthy features of Ket from a typological and areal perspective is its elaborate and highly complex verbal system. There are many different structural positions or slots within the Ket verb, the exact number of which is debated by Ket specialists (e.g., Werner (1997) assigns 18 such slots (14 prefix positions, a root, and three suffix positions), while Vajda (2003), on the other hand, considers there to be 10 such slots in the Ket verb (eight prefix position, root + one suffix position). Some examples of complex Ket verb forms include the following:

- | | |
|--------------------------------|--------------------------------|
| (6) <i>da-bágdèŋ-ú-γ-à-vet</i> | <i>d-us'-n-ba-γɔ-v-in'-tet</i> |
| 3FEM-drag- | 3M-INCORP-DER-1-PST- |
| 3NEUT.OBJ-DIR- | INAN-PST.PRF-hit |
| PRES-ITER | |
| 'she drags it often' | 'he has hit me' |
| (Vajda, 2003: 63) | (Werner, 1997: 156) |

Morphosyntactically, Ket stands out for its unusual predilection for multiple encoding of a single category, e.g., subject in verbs or plurality in nouns. Examples of this kind of redundant encoding may be seen in the following examples.

- | | |
|---------------------|--------------------------------------|
| (7) Ket | |
| <i>d-dəŋ-b-it-n</i> | ¹ <i>qa-ŋ-s-eŋ-naŋal'</i> |
| 1-1PL-INAN- | chief..-PL-..chief-PL-PL:ABL |
| transport-PL | |
| 'we transport it' | 'from the chiefs' |
| (Shabaev, 1987; | |
| Werner, 1995) | |
| | < ⁴ qa[:s] 'chief' |

A number of proposals have been offered on the possible wider genetic affiliation of the Yeniseic languages, including connections with Burushaski, Sino-Tibetan, and Northeast Caucasian languages of Eurasia, as well as Athabaskan (Na-Dene) languages of North America. To date, only the latter proposal has met with any positive reactions among specialists.

Bibliography

- Anderson G D S (1996). 'Ket tonal alternations in absolute plural.' In Aronson H I (ed.) *Non-Slavic languages 8: linguistic studies, new series*. Chicago: Chicago Linguistic Society. 1–18.
- Anderson G D S (2003). 'Yeniseic languages from a Siberian areal perspective.' In Vajda E & Anderson G D S (eds.) *Studia Yeniseica: typological studies on Yeniseic in honor of Heinrich Werner, Sprachtypologie und Universalienforschung* 56(1/2). München/Bremen: Akademie Verlag. 12–39.
- Belimov E I (1991). *Ketskij sintaksis*. Novosibirsk.
- Bibikova V S (1971). 'Grammaticheskie priznaki imeni prilagatel'nogo v ketskom jazyke.' *Jazyki i toponimija Sibiri* 4, 27–36.
- Bouda K (1957). 'Die Sprache der Jenissejer: Genealogische und morphologische Untersuchungen.' *Anthropos* 52, 65–134.
- Castrén M A (1858). *Versuch einer jenissej-ostjakischen und kottischen Sprachlehre nebst Wörterverzeichnissen aus den genannten Sprachen*. St. Petersburg.
- Comrie B (1982). 'Verb agreement in Ket.' In Aronson H I (ed.) *Folia Slavica* 5.1–3. *Papers from the Second Conference on the non-Slavic Languages of the USSR*. Columbus, Ohio: Slavica. 115–127.
- Denning R F (1976). 'Palatalizatsija soglasnykh v ketskom jazyke.' *Jazyki i toponimija Sibiri* 6, 85–89.
- Donner K (1930). 'Über die Jenissei-Ostjaken und ihre Sprache.' *Journal de la Société Finno-Ougrienne* 44, 1–32.
- Dul'zon A P (1961). 'Slovarnye materialy XVIII v. po enisejskim narechijam.' *Uchenye zapiski Tomskogo pedagogicheskogo instituta* 19(2), 152–189.
- Dul'zon A P (1966). *Ketskije skazki*. Tomsk.
- Dul'zon A P (1968). *Ketskij jazyk*. Tomsk: Izdatel'stvo tomskogo universiteta.
- Gajer R S (1981). 'Skazki i bytovye teksty.' In Porotova T I (ed.) *Skazki narodov sibirskogo Severa* 4. Tomsk. 27–40.
- Grishina N M (1981). 'Slozhnopodchinennye predlozhenija.' In Porotova T I (ed.) *Skazki narodov sibirskogo Severa* 4. Tomsk. 92–121.
- Karger N K (1937). 'Ketskij jazyk.' In *Jazyki i pis'mennost' narodov Severa III*. Leningrad: Krejnovich, È. A. 223–238.
- Khelimskij E A (1986). 'Arkhivnye materialy XVIII v. po enisejskim jazykam.' In Ja Skorik P (ed.) *Paleoaziatskie jazyki*. Leningrad: Nauka. 179–212.
- Kostjakov M M (1979). 'Vremja raskhozhdenija ketskogo i kottskogo jazykov po dannym leksikostatistiki.' In Ubrjatova E I (ed.) *Voprosy stroja enisejskikh jazykov*. Novosibirsk: SO Akademija Nauk SSSR. 118–127.
- Kostjakov M M (1986). 'Slozhnoe slovo i slovosochetanie v ketskom jazyke.' In Vall M N (ed.) *Issledovanija po grammatike i leksike enisejskikh jazykov*. Novosibirsk: Nauka. 59–69.
- Krejnovich E A (1968a). *Glagol ketskogo jazyka*. Leningrad: Nauka.
- Krejnovich E A (1968b). 'Sposoby dejstvija v glagole ketskogo jazyka.' In Ivanov V V et al. (eds.) *Ketskij sbornik. Lingvistika*. Moscow: Moskovskij Gosudarstvennyj Universitet. 75–138.
- Krejnovich E A (1969). 'Medvezhij prazdnik u ketov.' In Ivanov V V et al. (eds.) *Ketskij sbornik. Mifologija, etnografija, teksty*. Moscow: Moskovskij Gosudarstvennyj Universitet. 6–112.
- Krivosnogov V P (1998). *Kety na poroge III tysjacheletija*. Krasnojarsk: Krasnojarskij pedagogicheskij universitet.
- Maksunova Z (2003). 'Compounding and incorporation in Ket (central dialect).' In Vajda E J & Anderson G D S (eds.) *Studia Yeniseica: typological studies on Yeniseic in honor of Heinrich Werner, Sprachtypologie und Universalienforschung* 56(1/2), 123–132.
- Minaeva V P (2003). 'Russian grammatical interference in Ket.' In Vajda E J & Anderson G D S (eds.) *Studia Yeniseica: Typological Studies on Yeniseic in honor of*

- Heinrich Werner, *Sprachtypologie und Universalienforschung* 56(1/2), 40–54.
- Porotova T I (1990). *Kategorija mnozhestvennosti v enisejskikh jazykakh*. Tomsk: Izdatel'stvo tomskogo universiteta.
- Shabaev V G (1987). 'Dvojnaja markirovka subekta v sostave finitnykh form ketskogo glagola.' In Morev J A & Denning R F (eds.) *Jazyki narodov SSSR*. Novosibirsk: Novosibirsk State University Press. 100–106.
- Vajda E J (1999). *Languages of the World 15. Ket prosodic phonology*. Munich: Lincom.
- Vajda E J (2001a). *Yeniseian peoples and languages*. Surrey: Curzon Press.
- Vajda E J (2001b). 'The role of position class in Ket verb morphophonology.' *Word* 52(3), 369–436.
- Vajda E J (2002). 'The origin of phonemic tone in Ket.' *Chicago Linguistic Society* 37(2), 290–305.
- Vajda E J (2003). 'Ket verb structure in typological perspective.' In Vajda E & Anderson G D S (eds.) *Studia Yeniseica: typological studies on Yeniseic in honor of Heinrich Werner, Sprachtypologie und Universalienforschung* 56(1/2), 55–92.
- Verner G K (1991). *Sravnitel'naja fonetika enisejskikh jazykov*. Taganrog: Taganrogskij pedagogicheskij institut.
- Werner H K (1972). 'Die Tonität der gegenwärtigen Jenissei-Sprachen.' *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung* 25(1–2), 111–125.
- Werner H K (1995). *Zur Typologie der Jenissei-Sprachen*. Wiesbaden: Harrassowitz.
- Werner H K (1997a). *Die ketische Sprache*. Tunguso-Sibirica 3. Wiesbaden: Harrassowitz.
- Werner H K (1997b). *Abriß der kottischen Grammatik*. Tunguso-Sibirica 4. Wiesbaden: Harrassowitz.
- Werner H K (1997c). *Das Jugische (Sym-Ketische)*. Veröffentlichungen der Societas Uralo-Altaica 50. Wiesbaden: Harrassowitz.

Khasi

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

The Khasi are a group of Mon-Khmer speakers living predominantly in the Khasi and Jaintia Hills region of Meghalaya state in northeastern India, with a smaller number in Assam, West Bengal, and Manipur states. In some sources, the Khasi have been called Khuchia. The vast majority of Khasi people (ca. 90%) live in India, with a further 10% living across the border in Bangladesh. Khasi [KHI] is the only language of the Mon-Khmer group of Austroasiatic spoken this far to the west.

Traditionally the Khasi are divided into a number of 'dialect' groups, but perhaps it is more sound from a linguistic perspective to speak of a small group of related languages sometimes labeled Khasic or Khasian. These 'dialects' or closely related languages include the following (Grimes, 2000; Parkin, 1991):

- (1) Khasic languages
 - a. Amwi [AML]
 - b. Bhoi
 - c. Lyngngam
 - d. Pnar (aka Synteng or Jaintia) [PBV]
 - e. Khynriam or Cherrapunji/Standard Khasi [KHI]
 - f. War

Of these dialects/languages, Lyngngam is most linguistically distant from the standard Khasi dialect, while Pnar is the closest. Amwi is also quite distant

from Standard Khasi. Lyngngam may include a linguistically Khasified Garo element (a Tibeto-Burman language), while War and Bhoi may include assimilated Mikir (Tibeto-Burman) elements. The Pnar (Synteng/Jaintia) ruled a kingdom in the region from at least 1500 to 1835, when it was disbanded by the British colonial authorities (Parkin, 1991: 58). Further Khasic varieties include Lakadong and Mynnar. Other local Khasic varieties may (in fact probably do) exist, and there is also considerable microlevel variation.

According to figures from the India Missions Association in 1997, there were just under 1 million total Khasi speakers, including all the above mentioned dialects/languages (the actual estimated figure is 950 000). Khasi is a literary language and a language of media and government in Meghalaya. There are even radio and television broadcasts in the Standard Khasi language. Phonologically, Khasi exhibits some areally and typologically atypical initial clusters, e.g., [bt], [ks], [kt], [ktH], and so on.

- (2) *bta* 'wash/besmead face' *ksew* 'dog' *kti* 'hand'
kt^báw 'grandfather'

Syntactically, Khasi is SVO, while other Khasic languages show different basic word orders as well as many other different features.

- (3) *Khasi*
phi-m ?iithu? ya na
you-NEG recognize OBJ I
'don't you recognize me?'
(Rabel, 1961: 61)

Morphosyntactically, Khasi is characterized by use of gender markers and a system of personal verb inflection, albeit within a predominantly isolating structure.

- (4) Standard Khasi
u khinna? u-m bam
 DET.M boy MASC-NEG eat
 'the boy doesn't eat'
 (Nagaraja, 1993: 5)

In addition, there is evidence of a now (mainly) covert noun-class system that manifests itself in the form of lexicalized prefixes in noun stems. This system of gender classifiers is highly marked for Austroasiatic. This and the unusual phonology of Khasi set it apart from its sister languages spoken to the east.

In terms of verbal derivational morphology, Khasi, like its sister languages to the east, makes use of a causative prefix consisting of a labial consonant in various allomorphic realizations.

- | | | |
|--|--|--|
| (5) i. <u>Khasi</u> <i>ph-rung</i> 'penetrate' | <u>Khasi</u> <i>ph-láit</i> 'clear away' | <u>Khasi</u> <i>b-ta</i> 'wash/ besmear face' |
| < <i>rung</i> 'enter' (Henderson, 1976b: 487) | < <i>láiit</i> 'be free' (Henderson, 1976b: 487) | (Henderson, 1976b: 487) |

Negative occurs in the form of either an enclitic to a subject pronoun or gender agreement marker or proclitic to the verb stem, depending of the tense/aspect value of the clause. Note also the presence of fused subject+tense forms in Standard Khasi as well.

- (6) Standard Khasi
phi-m ?iithu? ya na *nan ?m-tho?*
 you-NEG recognize OBJ I I.FUT NEG-write
 'don't you recognize me?' 'I'm not writing'
 (Rabel, 1961: 61)

Standard Khasi
u khinna? u-m bam
 DET.M boy MASC-NEG eat
 'the boy doesn't eat'
 (Nagaraja, 1993: 5)

In Bhoi, another language of the Khasic subgroup of Mon-Khmer, the negative has a different phonological shape and occurs between the lexical verb and a postposed gender/agreement marker.

- (7) Bhoi
u khanna? bam re u
 DET.M boy eat NEG MASC
 'the boy doesn't eat'
 (Nagaraja, 1993: 5)

Bound aspectual or tense morphemes are rare in Eastern Austroasiatic. There is, however, a

quasi-bound suffixal past tense marker in *-la?* and a future in *-di?* in Lyngngam, a language of the Khasic subgroup.

- (8) Lyngngam
brə kyu di-la? liŋba la?tap
 man 3PL go-PAST through forest
 'the men went through the forest'
 (Nagaraja, 1996: 43)
- nə dənmi nə di-la? tu dənmi-di?*
 I go.NPAST I go-PAST he go.NPAST-FUT
 'I go' 'I went' 'he will go'
 (Nagaraja, 1996: 44)
- mi binməŋ-di? mi baŋ-la?*
 you eat.NPAST-FUT you eat-PAST
 'you will eat' 'you ate'
 (Nagaraja, 1996: 44)

In terms of nominal derivation, Khasi, like virtually all Austroasiatic languages, may derive deverbal nominals through a process of *-n-* infixation. Note that sometimes the derived noun reflects a more archaic phonological form than the verb stem it (historically) derives from, e.g., the preservation of initial *s-* in the word for 'wing' while the corresponding verb stem 'fly' has shifted this to *h-*.

- (9) Khasi
shnong *sner*
 'village' < *shong* 'feather, wing' < *her* 'fly'
 'sit, dwell'
 (Henderson, 1976b: 517–518)

Like many Austroasiatic subgroups (Anderson and Zide, 2002), Khasic languages show irregular correspondences in the free-forms of nouns, while the corresponding 'underlying' (usually CVC) roots are clearly cognate across the subgroup. Note the following forms in this regard:

- (10) Irregular Khasic correspondences
- | <u>Khasi</u> | <u>Lyngngam</u> | <u>Synteng</u> | <u>Amwi</u> | <u>Lakadong</u> |
|--------------|-----------------|----------------|----------------|-----------------|
| <i>ksew</i> | <i>ksu:/su:</i> | <i>ksaw</i> | <i>ksiá</i> | <i>ksaw</i> |
| | | ~ <i>kswa</i> | | |
| <i>sim</i> | <i>sim</i> | <i>sim</i> | | |
| <i>khmat</i> | <i>kh'mat</i> | <i>khmat</i> | <i>ma:t</i> | <i>ma:t</i> |
| <i>khmut</i> | leo- 'mut | <i>khmut</i> | <i>mur-koŋ</i> | <i>mur-koŋ</i> |
| | <u>Mynnar</u> | <u>War</u> | gloss | |
| | <i>ksow</i> | <i>ksià</i> | 'dog' | |
| | <i>ksem</i> | <i>ksem</i> | 'bird' | |
| | | <i>ma:t</i> | 'eye' | |
| | | <i>myrkoŋ</i> | 'nose' | |

(Fournier, 1974: 86–92)

That elements like *k-/kh-* are historically prefixes in Khasi is attested to by such facts as the following alternations. The original root forms appear as CVC 'combining form' elements in compounds.

- (11) **Khasi**
kti but *tiipden* 'middle finger.' (Rabel, 1961: 44)
khmat but *matli?* 'white of eye.'
 also *ʔiimat* 'eye' < see-eye/face (Rabel, 1961: 149)
khnaay 'mouse. rat' but *naaysaaw* 'small red
 hill mouse'

Note in this regard also the following alternations:

- (12) **Khasi**
kpa, *kmi(e)* 'father, mother' (non-vocative) vs.
ʔii paa ʔii mey address term used by children to
 parents 'da ddy/mommy' (Rabel, 1961: 49)

Bibliography

- Anderson G D S & Zide N H (2002). 'Issues in Proto-Munda and Proto-Austroasiatic nominal derivation: the bimoraic constraint' In Macken M A (ed.) *Papers from the 10th annual meeting of the Southeast Asian Linguistics Society*. Tempe, AZ: Arizona State University. 55–74. [South East Asian Studies Program, Monograph Series Press.]
- Bars Rev E (1973). *Khasi-English dictionary*. Shillong: Don Bosco.
- Blah U E (1970). *Khasi self teacher*. Shillong: Chapala Book Stall.
- Daladier A (2002). 'Definiteness in Amwi: grammaticalization and syntax.' In Zribi H A & Daladier A (eds.) *Recherches Linguistiques de Vincennes* 31. Paris: Press Universitaires de Vincennes. 61–78.
- Fournier A (1974). 'Les Khasi, une population mon-khmer de l'Inde.' *Asie du Sud-Est et Monde Insulindien* 5(1), 79–96.
- Gabelentz H C von der (1858). 'Grammatik und Wörterbuch der Khassia-Sprache.' *Verhandlungen der königlichen Gesellschaft der Wissenschaften zu Leipzig, Philol.-historische Klasse* 10, 1–66.
- Grimes B (ed.) (2000). *The Ethnologue* (14th edn.). Dallas: SIL.
- Henderson E J A (1965). 'Final -k in Khasi: a secondary phonological pattern.' In Milner G B & Henderson E J A (eds.) *Indo-Pacific linguistic studies*. Amsterdam: North Holland. 459–466.
- Henderson E J A (1966). 'Khasi and the l-clusters in seventeenth century Tonkinese.' In Ba Shin U, Boisselier J & Griswold A B (eds.) *Essays offered to G. B. Luce*. (Artibus Asiae, supp. xxiii). Ascona. 139–150.
- Henderson E J A (1967). 'Vowel length and vowel quality in Khasi.' *Bulletin of the School of Oriental and African Studies* 30, 564–588.
- Henderson E J A (1976a). 'Khasi Initial Clusters.' In Jenner *et al.* (eds.). 523–538.
- Henderson E J A (1976b). 'Vestiges of Morphology in Modern Standard Khasi.' In Jenner *et al.* (eds.). 477–522.
- Jenner P N, Thompson L C & Starosta S (eds.) (1976). *Austroasiatic studies*. Honolulu: University of Hawaii. [Oceanic Linguistics, Special Publication, No. 13.]
- Nagaraja K S (1979). 'Contraction of Khasi nouns in compounds.' *Indian Linguistics* 40, 18–23.
- Nagaraja K S (1984a). 'Compounding in Khasi.' *Bulletin of the Deccan College Research Institute* 43, 79–90.
- Nagaraja K S (1984b). 'Reduplication in Khasi.' *Indo-Iranian Journal* 27(3), 189–200.
- Nagaraja K S (1985). *Khasi: a descriptive analysis*. Pune, Maharashtra, India: Deccan College.
- Nagaraja K S (1993). 'Khasi dialects: a typological consideration.' *Mon-Khmer Studies* XXIII, 1–10.
- Nagaraja K S (1996). 'The status of Lyngngam.' *Mon-Khmer Studies* 26, 37–50.
- Parkin R (1991). *A guide to Austroasiatic speakers and their languages*. Honolulu: University of Hawaii. [Oceanic Linguistics Special Publication 23.]
- Rabel L (1961). *Khasi: a language of Assam*. Baton Rouge: Louisiana State University Press.
- Roberts H (1891). *A grammar of the Khassi language*. London: Kegan Paul, Trench, Trübner and Co.
- Schmidt Pater W (1904). 'Grundzüge einer Lautlehre der Khasi-Sprache.' *Transactions of Royal Bavarian Academy of Sciences* 22(3), 657–810.
- Sharma H S (1999). 'A comparison between Khasi and Manipuri word order.' *Linguistics of the Tibeto-Burman Area* 22(1), 139–148.
- Subbarao K V & Tansen G M (2003). 'Wh-question formation in Khasi.' In Kansakar T R & Turin M (eds.) *Themes in Himalayan languages*. Heidelberg: South Asia Institute/Kathmandu: Tribhuvan University. 197–218.
- Weidert A K (1975). *I tkong Amwi. Deskriptive Analyse eines Wardialekts des Khasi*. Wiesbaden: Harrassowitz.

Khmer

M Minegishi, Tokyo University of Foreign Studies, Tokyo, Japan

© 2006 Elsevier Ltd. All rights reserved.

Khmer (Cambodian)

In the Kingdom of Cambodia, most of the population of 10 716 000 (1998 UN) are considered speakers of Khmer. Its dialectal varieties are spoken by around

1.3 million ethnic Khmer people in the northeastern and eastern provinces of Thailand. The former variety is called Northern Khmer, or sometimes Surin Khmer. Another variety is spoken by more than one million people of the Khmer ethnic group, called Lower Khmer, in southern Vietnam.

Khmer is one of the major languages of Mon-Khmer subgroup of the Austroasiatic language family. It is a typical Mon-Khmer language in that phonemically its

native words are either monosyllabic or disyllabic; it has no tonal distinction and is of the isolating type syntactically.

Script and Written Records

Khmer script is one of the oldest scripts in mainland Southeast Asia that originate in South India. Archeology has shown that the communication between Southeast Asia and India dates back to the beginning of the Christian era. Khmer legend, according to the Chinese document, says the local queen married the prince from India, and they became the founders of the Khmer kingdom, which suggests the existence of local matrimonial authority influenced by the Indian civilization, such as Hinduism.

The oldest Khmer inscription in old Khmer dates from 611 C.E., in addition to which there are also undated inscriptions, or presumably older ones written in Sanskrit. The inscriptions spread not only in Cambodia, but also in parts of Thailand and southern Vietnam, which suggests that the ethnic group was formerly more widespread than it is in the present and once exerted a strong cultural influence over the area.

Phonology and Phonetics

A Khmer native word is either monosyllabic or disyllabic. In a disyllabic word, a minor syllable precedes a major syllable. In a minor syllable, the inventory of possible vowels is smaller than in a major syllable. A major syllable is pronounced with stress when preceded by a minor syllable. Using the abbreviations C for a consonant, V for a vowel, F for a syllable final consonant, 'r' for a liquid and parentheses for an optional element, a minor syllable can be either CVF or C(r)V, where only nasals can appear as F.

Likewise, a Khmer major syllable can be illustrated as C (C) V F. Vowels are either long or short. The long and short contrast is also found among diphthongs. A short vowel occurs only in a checked syllable: it must be followed by a syllable-final consonant, whereas a long vowel can occur both in an open and checked syllable.

Consonants

Consonants in the syllable-initial position are given in Table 1, where the IPA symbols, when necessary, are given in the brackets.

Of these consonants, only nasals /m, n, ɲ ɳ/, unreleased stops /p̚, t̚, c̚, k̚/, a glottal stop /ʔ/, fricatives /v, y, h/ and a liquid /l/ can occur in the word-final position as well. /r/ in the word-final position, once pronounced, is lost except in the Northern Khmer dialect in Thailand. /f/ appears only in loanwords.

Table 1 Khmer consonants

| | | | | |
|-------|--------|----------|----|---|
| p | t | c [tɕ] | k | ʔ |
| ph | th | ch [tɕh] | kh | |
| b[β] | d[d] | | | |
| m | n | ɲ | ɳ | |
| | r l | | | |
| (f) v | s y[j] | | | h |

Table 2 Khmer long vowels and diphthongs

| Phonemic | Orthography |
|----------|-------------|
| i: | ī-2 |
| e: | e-2 |
| ɛ: | e-1 |
| æ: | æ-2 |
| æ: | æ-1 |
| u: | ū-2 |
| ɔ: | ō-2 |
| ə: | ē-1 |
| aə | ā-1 |
| iə | ā-2 |
| a: | ā-1 |
| u: | ū-2 |
| ɔu | o-2 |
| o: | ū-1 |
| ɔ: | a-2 |
| ɔ: | a-1 |
| ao | o-1 |

The distinction between voiceless unaspirated and voiceless aspirated stops can be found only in the syllable initial or intervocalic position. Voiceless aspirated consonants could be further analyzed as consonant clusters /p/ + /h/, /t/ + /h/, /c/ + /h/, /k/ + /h/, as there are some words that have an infix between the first stop and the second fricative /h/. Characteristic of the Khmer consonants is that Khmer allows a variety of two-consonant combinations in the syllable initial position of a major syllable. Regarding the above aspirated stops as consonant clusters, 84 combinations in total are possible.

Vowels

Modern Khmer has a complicated vowel system. Although several dictionaries have been published for Khmer, there is no consensus as to the vowel phonemic system. As a result, almost every dictionary has its own phonemic transcription. The main reason for the discrepancy is that some assume the existence of resister contrast, i.e., contrast between 'breathy' and 'clear' phonation type, but others do not.

Table 2 shows the standard Khmer long vowels and diphthongs in phonetic transcription with the transliteration of orthography of Indic origin. The register

Table 3 Khmer short vowels and diphthongs

| | | |
|----|---|---|
| i | u | u |
| e | ə | o |
| ɛ | | ʊ |
| ɛə | a | ɔ |

contrast, which had been regarded as phonemic, has been lost in the dialects in Cambodian territory, and diphthongization has occurred in compensation. See Minegishi (1985) for details of phonetic values of Khmer dialects. According to Wayland and Jongman (2003), the remnant of phonation contrast is observed in the dialect of eastern Thailand.

Note that, “æ, ɪ, ə” are transliterations for Khmer original scripts, which do not exist in the ordinary Indic script system. A ‘1’ following the transliteration means that the consonant preceding the vowel symbol is of voiceless group; ‘2’ indicates the preceding consonant of the voiced group. In addition, there are diphthongs /uə, əə, uə/. In total, there are 12 long vowels and eight diphthongs. This complexity is attributable to the loss of voiceless and voiced contrast in the syllable-initial position for stops and successive divergence of vowels. See ‘Historical Phonology’ below for details. Table 3 shows the Khmer short vowels and diphthongs.

Historical Phonology

The phonemic reconstruction by Sakamoto based on Khmer inscriptions has established the old Khmer vowel phonemes as */i, e, ɛ, a, aa, (u), ə, u, o, ɔ, ʊ, ɐ, iə, uə/.

By the above reconstruction and modern orthography, diachronic changes in their phonological systems can be internally reconstructed as follows.

Formerly, Khmer had a phonemic contrast, e.g., /*kaa/ and /*gaa/ as its orthography shows, where the difference is between voiceless and voiced consonants. Later, the vowel following the voiced consonant changed its quality; /*kaa/ and /*gèa/, where phonemic contrast between the consonants still existed and the difference in the vowel register, i.e., ‘clear’ versus ‘breathy’ phonation type respectively, was irrelevant. Later on, the voiceless and voiced distinction in stops was lost and the difference in the voice quality in turn carried the phonemic contrast; /*kaa/ and /*kèa/. In the present, the voice quality is no longer phonemic; instead vowel articulation is relevant; /ka:/ and /kiə/. The divergence of vowels is well preserved in the standard Khmer around Phnom Penh area, but in the rest of the country, vowels have merged again to simplify the vowel system. As a result, most of the dialects have only /i:, e:/ as long front vowels, etc. Northern

Khmer, conversely, has retained most of the vowel contrast as monophthongs.

Tonal Contrast

As is usual in Mon-Khmer languages, Khmer does not have tones. The only exception, however, is the colloquial style of Phnom Penh dialect, which has acquired a tonal contrast, a level tone versus a raising-falling one: the latter is a compensation for the phonemic change /r/ into /h/, and the successive loss of /h/.

Morphology

Khmer, although syntactically of an isolating type, has a large number of derivational prefixes and infixes, which have been fossilized and are no longer productive in word formation. A word may have either a prefix or infix, but not both. Thus, native words are either monosyllabic or disyllabic. Like other Mon-Khmer languages, Khmer has no suffix. It has prefixes and infixes for causativization, specification, nominalization, intransitivization, repetition, etc., and infixes representing instrument, agent, result, object, etc., of an action.

Grammar

Khmer is an isolating language with no inflection in verbs, nor case marking in nouns. As a result, classifying word classes must be done by means of their distribution and class meaning. Noun modifiers follow the noun, verb modifiers follow the verb.

Khmer’s main word classes are as follows, although further classification considering the syntactic distribution is possible: nouns, numerals, classifiers, demonstratives, pronouns, verbs, preverbs, adverbs, expressives, conjunctions, and final particles.

Of these, classifiers are few in number and rarely used except for counting persons, animals, or books. There are two demonstratives. Pronouns are a subclass of nouns, most of which are also used as nouns. Along with titles and kinship terms, choice of pronouns shows relative social positions. Verbs can be further classified as active and stative (adjectival) verbs. Preverbs may precede verbs adding modal meaning, such as ‘may, must,’ etc., to them. Adverbs may follow a verb. Expressives are a subclass of adverbs, describing noises, shapes, movements, emotions, etc. Final particles may be in the sentence-final position to denote the intentions and emotions of the speaker, etc.

The basic word order is Subject+Verb+(Object). A modified noun (head) is followed by a modifier. Nouns and stative verbs can be used as modifiers. Prepositions

(or a noun grammaticalized as a preposition) precede a noun.

The typical noun phrase can be described as follows where optional elements are in parentheses.

Noun+(Verb)+(Numerals)+(Classifier)+(Demonstrative)

In case a clause is used for the noun modifier, the word order is as follows.

Noun+(Relative clause marker+Clause)+(Demonstrative)

A verb may be followed by a noun to form a verb phrase. Several verbs, sometimes with a noun inserted in between, form a serial verb construction, or verb serialization without any change in verbal forms, such as V(N)V(N), etc. In a serial verb construction, two or more verbs may be in various semantic relations, such as an action and its direction, an action and its objective, successive actions, an action and its result, an action and its manner, or an action and its means, etc.

Vocabulary

As one of the earliest languages in Southeast Asia that has accepted Indianization, earlier Hinduism, and later Theravada Buddhism, Khmer has borrowed Sanskrit and Pali loans, especially for religion, administrative, and other cultural vocabulary. It also has exerted a huge influence over adjacent Thai, which in turn borrowed a large number of Khmer words, such as honorific vocabulary used for the royal family. As a result of long-term contact with Thai, Khmer also borrowed many words from Thai.

Bibliography

- Chantharupanth T & Phromjagarin C (1978). *Surin Khmer-English dictionary*. Bangkok: Institute of Indigenous Language Research, Chulalongkorn University Press.
- Diffloth G (2003). 'Austroasiatic language.' *Encyclopedia Britannica 2003* (DVD edn.). Encyclopedia Britannica.

- Headley R K Jr, Chhor K, Lim L K, Lim H K & Chun C (1977). *Cambodian-English dictionary*. Washington DC: The Catholic University of America Press.
- Henderson E J (1952). 'The main features of Cambodian pronunciation.' *Bulletin of the School of Oriental and African Studies* 14(1), 149-174.
- Huffman F (1977). *Cambodian reader and glossary (Yale linguistic series)*. New Haven and London: Yale University Press.
- Huffman F & Proum I (1977). *English-Khmer dictionary*. New Haven, CT: Yale University Press.
- Institute bouddhique (1967). *Dictionnaire cambodgien*. Phnom Penh.
- Jacob J M (1968). *An introduction to Cambodian*. London: Oxford University Press.
- Jacob J M (1974). *A concise Cambodian-English dictionary*. London: Oxford University Press.
- Jenner P N (1974a). 'The development of the registers in Standard Khmer.' In Liem N D (ed.) *Southeast Asian linguistic studies I*. Canberra: Australian National University. 47-60.
- Jenner P N (1974b). 'Observations of the Surin dialect of Khmer.' In Liem N D (ed.) *Southeast Asian linguistic studies I*. Canberra: Australian National University. 61-73.
- Martin M A (1975). 'Le dialecte cambodgien parlé à Tatêy, massif des Cardamomes.' *Asie du Sud-est et Monde Insulindien* 6(4), 71-79.
- Maspéro G (1915). *Grammaire de la langue khmère (cambodgien)*. Paris: Imprimerie Nationale.
- Minegishi M (1985). 'Modern Khmer orthography and its dialects.' [in Japanese] *Gengo Kenkyu* 88.
- Sakamoto Y (1970a). 'i, i, ya, yā de khmer ancien.' *Journal of Southeast Asian Studies* 7, 4. 492-503.
- Sakamoto Y (1970b). 'e de khmer ancien.' *Journal of Asian and African Studies* 3, 28-43.
- Sakamoto Y (1971). 'Sur quelques voyelles de khmer ancien.' *Journal of Asian and African Studies* 4, 55-73.
- Sakamoto Y (1974). 'a et ā de khmer ancien.' *Journal of Asian and African Studies* 4, 75-100.
- Sakamoto Y (2001). *Cambodian-Japanese dictionary* (3 vols). Tokyo: ILCAA, Tokyo University of Foreign Studies.
- Wayland R & Jongman A (2003). 'Acoustic correlates of breathy and clear vowels: the case of Khmer.' *Journal of Phonetics* 31(2), 181-201.

Khoesaaan Languages

W H G Haacke, University of Namibia, Windhoek, Namibia

© 2006 Elsevier Ltd. All rights reserved.

The recognition of Khoisan (Khoesaaan) as one of the four language phyla in Africa received initial impetus mainly through the 'Macro-Khoisan Hypothesis' of

Joseph Greenberg, first published in 1950. This controversial phylum comprises his 'Click languages,' which formerly were known as 'Hottentot' and 'Bushman' languages, respectively. Dorothea Bleek (1929) had paved the way for their integration into one family by challenging the prevalent view that Hottentot as a Hamitic language had been influenced by Bushman, and by suggesting instead

Table 1 Southern African Khoesaaan languages

| Branch | Languages († = [virtually] extinct) |
|-----------------------------------|---|
| 1. Northern (<i>Ju</i>) | !Xung (DC), Jul'hoan, Xaull'e, !O!xung |
| 2. Southern (! <i>Ui-Taa</i>) | |
| 2.1 †! <i>Ui</i> | †!Xam, †!Aunī, † ≠ Khomanī, †!Xegwi |
| 2.2 <i>Taa</i> Isolate | !Xōō (DC), †Kakia ≠ Hōā |
| 3. Central (<i>Khoe</i>) | |
| 3.1. Khoekhoe | |
| 3.1.1. Northern/Namibian | Khoekhoegowab DC (= Nama, Damara, Haill'om, ≠ Ākhoe) |
| 3.1.2. †Southern/South African | †!Gora ('Korana'), †Xri ('Griqua'), †Cape Khoekhoe (DC) |
| 3.2. Kalahari Khoe | |
| 3.2.1. Western | Khwe, Buga, Ani (DC); Naro (DC); Gana, Gui, ≠ Haba (DC); |
| 3.2.2. Eastern | Shua, Ts'ixa, Danisi, Xaise, †Deti; Kua-Tsua (DC) |
| 4. | †Kwadi (Angola) |
| 5. | Sandawe (Tanzania) |

that Nama (as representative of Hottentot) was a Bushman language with Hamitic admixture. Bleek divided the Bushman languages into Northern, Central, and Southern groups. In essence, these divisions are still recognized today, although their validity is open to challenges (see Table 1).

The compounded name Khoisan was coined in 1928 by Leonhardt Schultze to signify the somato-racial relatedness of the Hottentots and Bushmen, with *khoi* ('human being,' correctly spelled *kho*) representing Hottentot, and *san* – the Nama designation for the Bushmen, meaning 'foragers' (correctly spelt *Sān* or *Saan*) – Bushman.

Greenberg distinguished South African Khoisan (with the major Northern, Central, and Southern branches) as opposed to the East African isolates Sandawe and Hadza (some 70 000 and 400 speakers, respectively). These views are not unanimously accepted, as genetic relatedness between the major linguistic branches cannot be proved satisfactorily. Although scholars remain divided on the issue of genetic relatedness, the term 'Khoesan' or 'Khoisan' (more correctly spelled 'Khoesaaan') is now widely used as a term of convenience to denote all non-Bantu and non-Cushitic click languages of Southern and Eastern Africa.

Only some 30 Khoesaaan languages still exist today, with the great majority of languages being extinct. With the possible exception of Khoekhoegowab in Namibia (formerly better known as 'Nama,'

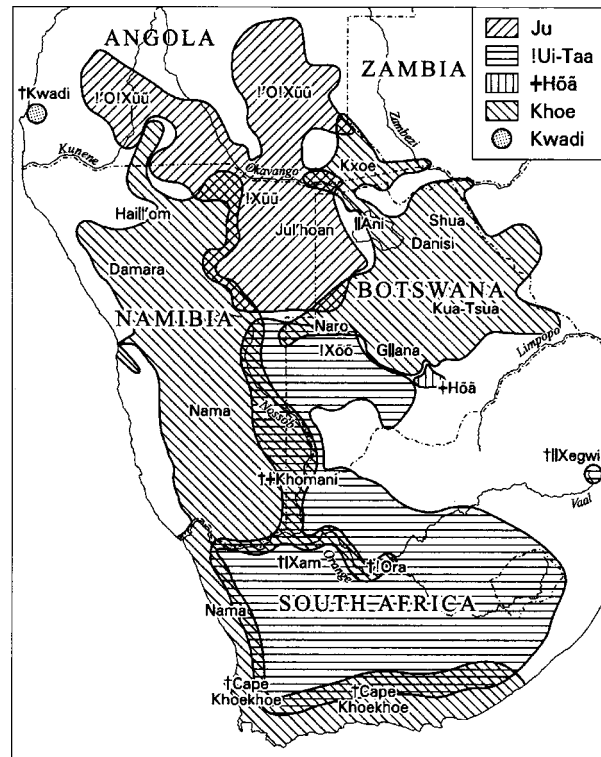


Figure 1 South African Khoesaaan languages (precolonial situation). (Note: '†' = language now extinct.) Map reproduced from Güldemann T & Vossen R (2000). 'Khoisan.' In Heine B & Nurse, D (eds.). *African languages: an introduction*. Cambridge: Cambridge University Press. 100, map 5.5.

and for classificatory purposes briefly referred to as 'Khoekhoe'), virtually all of these languages can be considered to be endangered.

While the Northern (*Ju*) and Southern (!*Ui-Taa*) branches with the isolate ≠Hōā are spoken by hunter-gatherers (Bushman/Saan), the languages of the Central branch (*Khoe*) are today spoken by Khoeid (Nama; !Gora/!Ora, Xri, and Cape Khoekhoe being extinct), Saaid (of especially the Kalahari Khoe branch), and Negroid (Damara) peoples. Linguistic and racial classifications of these groups are thus not coextensive. The following classification is largely based on the classifications of Köhler (1989) and of Güldemann and Vossen (2000). The reader is referred to the latter publication for more detailed information on Khoesaaan languages. For a classification of !Xung and Jul'hoan dialects (Northern Khoesaaan) see Snyman (1997), for Central Khoesaaan see Voßen (1997), and for dialects of Khoekhoe(gowab) Haacke *et al.* (1997). Several of the language names below represent dialect clusters (DCs). The iconic classificatory names *Ju*, !*Ui*, *Taa*, and *Khoe* mean 'human being' in their respective branches. The now extinct Kwadi was probably related to Namibian Khoekhoe.

Table 2 The 20 click variants of Khoekhoegowab

| Influx | Efflux | | | | |
|----------|--------------|----------------------|---------------------------|---------------------------|---------------------|
| | Glottal stop | Voiceless velar stop | Delayed glottal fricative | Voiceless velar affricate | Voiced nasalization |
| Dental | [ʔ] | [g] | [h [ʔh] | [kh [xʰ] | [h [ŋ] |
| Alveolar | ! [!ʔ] | !g [!]] | !h [!ʔh] | !kh [!xʰ] | !n [ŋ!] |
| Palatal | ≠ [≠ʔ] | ≠g [≠] | ≠h [≠ʔh] | ≠kh [≠xʰ] | ≠n [ŋ≠] |
| Lateral | [ʔ] | g [] | h [ʔh] | kh [xʰ] | n [ŋ] |

Although Sandawe does show evidence of affinity to Central Khoesaaan, *i.a.* also in its tonology, recent research by Bonny Sands (1998b) considers Hadza of Tanzania to be an isolate with no satisfactory evidence for a genetic relationship to Khoesaaan, despite lexical similarities and the use of clicks (Figure 1).

Before the Bantu diaspora, Khoesaaan peoples probably inhabited the entire Southern Africa up to the east coast of South Africa and into southern Angola. Displacement and absorption led to a drastic reduction of territorial domains, which was further aggravated by the arrival of European colonizers in the seventeenth century. Social marginalization still determines the life of all groups considered to be Saan. Khoekhoegowab in Namibia is the only Khoesaaan language that is officially recognized for language-planning purposes and is a major subject at university level. Demographic figures are largely based on estimates, totalling over 200 000 for Southern African Khoesaaan. Khoekhoe(gowab), with 175 554 speakers (1991 census of Namibia), represents by far the largest Khoesaaan speech community and constitutes some 12.5% of the Namibian population.

The most conspicuous phonological characteristic of Khoesaaan languages is the use of click consonants (Table 2). Clicks consist of an influx and an efflux phase. The influx (basic click) produces the actual clicking sound and is produced without pulmonic airstream. There are five influx variants: ☉ (bilabial), | (dental), ! (alveolar), ≠ (palatal), and || (lateral). Each of these influxes then combines with a specific number of effluxes, depending on the language. This efflux constitutes the resumption of the pulmonic egressive airstream, and its nature depends on the manner of release of the posterior, velaric (and at times glottalic) closure. The bilabial ‘kiss-click’ ☉ is manifest only in Southern Khoesaaan. The number of effluxes can vary from five in Khoekhoe(gowab) to 16 in !Xōō, yielding a total of 83 click variants in the latter language, with a total of 117 consonants (cf. Traill, 1985) – a possible world record.

The tonology of Central Khoesaaan languages is typologically most akin to those of Southeast Asian languages on account of perturbational (sandhi)

Table 3 The six main citation melodies of Khoekhoegowab and their sandhi correlates (as recorded by Eliphas Eiseb)

| Citation | Sandhi | Gloss |
|----------|--------|--|
| !om̩s | !om̩s | to butt |
| !om̩s | !om̩s | female genitals, udder |
| !om̩s | !om̩s | to force escape from burrow (of: aardvark) |
| !om̩s | !om̩s | to coagulate; to remove thorn with aid of utensil |
| !om̩s | !om̩s | fist |
| !om̩s | !om̩s | pollard |

processes and the interaction of tonal and segmental phonemes, with depressor consonants triggering tonogenesis (development of contrastive pitch as compensation for the depletion of consonantal contrasts).

According to research thus far, available roots in Khoesaaan languages appear to be generally bimoraic. Roots of at least Central Khoesaaan languages are disyllabic, with syllable and mora being in isomorphic relation; hence tonal melodies (Table 3) consist of a sequence of two register (level) tonemes. The following minimal set illustrates the six main melodies of Khoekhoegowab in citation and sandhi form (Haacke, 1999). The second syllable of !om consists of a syllabic nasal [m]. Verbs are quoted here in their infinitival form with the third person fem. sg. pgn-marker s). Vowel qualities generally vary between oral and nasalized vowels; in addition, pharyngealized, laryngealized, and breathy vowels or their combinations are found in most non-Khoekhoe languages.

The most distinctive morphological characteristic of Central Khoesaaan languages is that they mark nouns for sex gender with postclitic person–gender–number markers, whereas Northern and Southern languages do not. This occurs most consistently in Khoekhoegowab, which marks nouns for person (third, as well as first or second), gender (masculine, feminine, neuter/common), and number (singular, dual, and plural). Non-Central languages have little inflectional morphology. Whereas Non-Central languages have SVO constituent order, Central languages have SOV order in the case of lexically specified NPs.

Bibliography

- Beach D M (1938). *The phonetics of the Hottentot language*. Cambridge: Heffer & Sons.
- Bleek D F (1929). *Comparative vocabularies of Bushman languages*. Cambridge: Cambridge University Press.
- Güldemann T & Vossen R (2000). 'Khoisan.' In Heine B & Nurse D (eds.) *African languages. An introduction*. Cambridge: Cambridge University Press.
- Haacke W H G (1999). 'The tonology of Khoekhoe (Nama/Damara).' In Voßen R (ed.) *Quellen zur Khoisan-Forschung* 16. Cologne: Köppe.
- Haacke W H G, Eiseb E & Namaseb L (1997). 'Internal and external relations of Khoekhoe dialects: a preliminary survey.' In Haacke W H G & Elderkin E D (eds.) (1997). 125–209.
- Haacke W H G & Elderkin E D (eds.) (1997). *Namibian languages: reports and papers*. Namibian African studies 4. Cologne: Köppe.
- Köhler O (1989). *Die Welt der Kxoe-Buschleute*. Berlin: Reimer.
- Sands B (1998a). *Research in Khoisan Studies 14: Eastern and Southern African Khoisan*. Cologne: Köppe.
- Sands B (1998b). 'The linguistic relationship between Hadza and Khoisan.' In Schladt M (ed.) *Language, identity, and conceptualization among the Khoisan*. Research in Khoisan Studies 15. Cologne: Köppe. 265–283.
- Schultze L (1928). *Zur Kenntnis des Körpers der Hottentotten und Buschmänner*. Jena: G. Fischer.
- Snyman J W (1975). *Žul'hōasi fonologie en woordeboek*. Cape Town: Balkema.
- Snyman J W (1997). 'A preliminary classification of the !Xŋ and Žul'hōasi dialects.' In Haacke W H G & Elderkin E D (eds.). 21–106.
- Traill A (1985). *Quellen zur Khoisan Forschung 1: Phonetic and phonological studies of !Xōō Bushman*. Hamburg: Buske.
- Traill A (n.d.) *Extinct South African Khoisan Languages* (audio-CD with booklet). Johannesburg: University of the Witwatersrand.
- Voßen R (1997). *Die Khoe-Sprachen. Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Research in Khoisan Studies 12. Cologne: Köppe.
- Westphal E (1971). 'The click languages of southern and eastern Africa.' In Sebeok T (ed.) *Linguistics in sub-Saharan Africa*. The Hague: Mouton. 367–420.

Khotanese

P O Skjærvø, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Khotanese, an Eastern Middle Iranian language spoken in the kingdom of Khotan in southwestern Xinjiang, is known from manuscripts on paper and wood found in the Khotan area, as well as in the caves at Dunhuang. The related language Tumshuqese, spoken in Kucha in northwestern Xinjiang, is much less well known. They were written in the southern and northern variants of Brahmi, respectively.

Three stages of Khotanese can be distinguished: Old, Middle, and Late, corresponding to texts from, roughly, the 5th–6th, 7th–8th, and 9th–10th centuries, up to the end of Buddhism in Khotan (ca. 1000).

Written remains consist mainly of Buddhist texts from all three periods, economical and legal documents from the Middle Khotanese period, and letters from the Late Khotanese period.

Khotanese and Tumshuqese constitute the north-eastern branch of the Iranian languages, in which Indo-Iran. *čw*, *fw* [tʃw, dʒw] became *ś* [ʃ], *ž* [ʒ]. This feature is only shared by modern-day Wakhi spoken in northeastern Afghanistan.

The phonology of Khotanese is of the Middle Iranian type in which, for instance, *č* [tʃ] and *ž* [dʒ],

spelled *c* and *j*, had become [ts] and [dz], spelled *tc*, and *js* in the Brahmi alphabet. There are two non-Indic vowel marks, transcribed as *-ä* ([*-ə*]?) and *-ei*.

Khotanese had retroflex consonants (*t̪*, *th̪*, *ḍ̪*, *ṇ̪*) in both indigenous words (in which, e.g., *ḍ̪* < *rd*, *ṇ̪* < *žn*) and in Indic loanwords. It is not clear whether the aspirated stops, *kh*, etc., were spirants [x], etc., or actually aspirated stops [kʰ], etc. It is possible that they were originally spirants, as in other Iranian languages (cf. Khot. *khara*- 'donkey,' Avestan *xara*-, Persian *xar*, etc.), and only later became stops, as suggested by the way Chinese was written in Brahmi in Khotan.

The non-Indic sound *z* was written *ys*. The voiced sibilants *ž* and *ž̌* were originally not distinguished in writing from *ś* and *ś̌*, but later various strategies for distinguishing them were invented (written double = unvoiced, single = voiced; a subscript curved line (transliterated as ') to indicate the voiced pair; e.g., *śārā* and *śśārā* [śārə] 'good,' but *śātā* [žādə], *śe* [že] 'second'); the line also probably indicated rhotasized vowels (*nei* [nei̠] 'nectar' < **nāžā*). There was a single-flap *r* and a trilled *r*.

While intervocalic voiced stops had already been lost in Old Khotanese, intervocalic *k* and *t* still remained in the oldest texts (phonemically *g* and *d*), and final *-i*, *-ä*, and *-u* were still distinct. The development from Old via Middle to Late Khotanese in the

main involved the loss of distinction between final short vowels and their loss after nasals and in some other positions (e.g., Okhot. *aysu* 'I' > MKhot. *aysä*, LKhot. *a*; OKhot. NOM/ACC *suhāvatānālu* 'utilities' > MKhot. *suhāvām* > LKhot. *suhāvau*); and the merger of *um* and *ām* (e.g., OKhot. GEN-DAT PL *rrumḍānu* 'the kings' > MKhot. *rrāmḍām* > LKhot. *rraudau* [approx. [rɔːdɔː]]).

The original six-case system is reduced in Middle Khotanese by the merger of the nominative and accusative (except in pronouns), and further reduced in Late Khotanese to a three- or two-case system. There are three genders, the neuter being of limited use.

The verbal system is of the Eastern Middle Iranian type. There are two stems, present and past (e.g., PRES *bar-*, PAST *buḍa-*; PRES *hām-* 'become,' PAST *hām-āta-*). Khotanese has preserved the Old Iranian moods (indicative, imperative, subjunctive, optative, injunctive), as well as active and middle. The past of intransitive verbs is of the common Iranian type (*buḍ-ā mā* 'carry.PAST.INTRANS-SING.MASCCOP.PRES.1ST.SING' = 'I was carried' > 'I rode'), while that of transitive verbs is based on an active past participle plus copula (*buḍ-e* < **bṛta-āh* 'carry.PAST.TRANS-SING.MASC [COP. 3RD.SING = Ø]' = 'he carried,' *buḍ-ātā* 'she carried'; *buḍ-aimā* < **bṛta-āh ahmi* [COP.PRES.1ST.SING] 'I [MASC] carried,' *buḍāndā* 'they carried,' *buḍ-āndā mā* [COP.PRES.1ST.PL] 'we carried').

Perfect/pluperfect and modal forms are formed from the past tense (e.g., perfect: *nei* [<*ne* + *ī*] *hvat-e štā balysā* 'NEG-EMPH.PART speak.PAST.TRANS-SING. MASC COP.3RD.SING Buddha-SING.NOM' = 'the Buddha has not at all said'; pluperfect: *cīyā rr-e bāysānd-ā vāt-ā* 'when king-SING.NOM awaken.PAST. INTRANS-MASC COP.PAST.INTR-3RD.SING-MASC' = 'when the king had awakened'; pluperfect optative: *ka nā va ysār-u gyast-a balys-a dāt-u hvat-āndā v-ī-ro ne gāvu vamas-īro* 'if they.ENCL.OBL

PARTICLE 1000. SING.NEUT lord.PL.NOM-ACC buddha.PL.-NOM-ACC law. SING.ACC speak.PAST.TRANS-PL.MASC/FEM COP-OPT-3RD.PL NEG at-all understand.OPT.3RD.PL' = 'even even if a thousand lord buddhas had spoken the law to them, they would not at all understand').

The 'potentialis' is formed with a past participle with the ending-*u* (SING ACC NEUT) and the verbs *yan*- 'to do' (active) and *hām*- 'become' (passive) and expresses possibility and completion of action (e.g., *ne hvat-a hām-āre* 'NEG speak.PAST.PART-PLUR.MASC become.PRES.-3RD.PL' = 'they cannot be said/expressed,' *ne hvat-u yan-īmā* 'NEG speak.PAST.PART. NEUT do.PRES-1ST.SING' = 'I cannot say (it),' *cīyā hvat-u yuḍ-āndā* 'when speak.PAST.PART.NEUT do.PAST.TRANS-3RD.PL = 'when they had spoken').

Tumshuqese also has the old augmented imperfect (e.g., *a-ch-i* 'PAST-go-3RD.SING' = 'he went, he has gone'), though the augment may be added only to monosyllabic forms (cf. *bar-i* 'he carried').

The lexicon contains numerous borrowings from Indic, both Middle Indic and Classical Sanskrit. In the Middle and Late Khotanese periods, we also find a small number of Chinese and Tibetan words.

Bibliography

- Bailey H W (1979). *Dictionary of Khotan Saka*. Cambridge: Cambridge University Press.
- Emmerick R E (1968). *Saka grammatical studies*. London: Oxford University Press.
- Emmerick R E (1989). 'Khotanese and Tumshuqese.' In Schmitt R (ed.) *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert. 204–229.
- Skjærvø P O (2002). *Khotanese manuscripts from Chinese Turkestan in The British Library: a complete catalogue with texts and translations*. Sims-Williams U Contributor. London: The British Library.

Kinyarwanda

A Kimenyi, California State University at Sacramento, Sacramento, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

General Background

Kinyarwanda, the national language of Rwanda is probably, after Swahili the second largest spoken language in the Bantu group. It is a sister dialect of Kirundi, the national language of Burundi, and Giha, another dialect spoken in Tanzania. Despite

the genocide that took place, taking the lives of more than 1 million Tutsi, it is spoken by perhaps more than 20 million people. Rwanda has approximately 9 million people right now, Burundi has approximately 7 million, but besides the Giha speakers there are also ethnic Banyarwanda in Southern Uganda in the Kigezi district known as Bafumbira. Other Kinyarwanda speakers are Banyamulenge in Southern Kivu, ethnic Banyarwanda in Masisi, and Rutshuro in Northern Kivu in the Democratic Republic of Congo. Kinyarwanda belongs to the interlacustrine (Great Lakes) Bantu languages.

Writing System

Although Kinyarwanda has both long vowels and short vowels as well as high tones or no tones on syllables, the official orthography does not mark vowel length and melody. Only the context can tell the reader which word was meant. Written texts are thus ambiguous even to native speakers. Thus, the written word *gusura* can stand either for (*gusura*) ‘to fart’ or (*gusuura*) ‘to visit’, and *gutaka* can stand for either (*gutaka*) ‘to scream’ or (*gutaaka*) ‘to decorate’; *ino* can stand for either (*ino*) ‘toe’ or (*inó*) ‘here’, *inda* can stand for (*inda*) ‘stomach’ or (*indá*) ‘louse’, *umuryango* can stand for (*umuryaango*) ‘family’ or (*umuryáango*) ‘door’, and *ikirere* can stand for (*ikireere*) ‘banana leaf’ or (*ikiréeré*) ‘air space’. Even though the sound ‘p’ has been lost and is found only in onomatopoeic words and loan words, the aspirated voiceless velar fricative ‘h’ is spelled as ‘p’ after the bilabial nasal ‘m’, as shown in the examples *impuha* (*imhuuha*) ‘rumors’, *impamvu* (*imhaámvu*) ‘cause/reason’. The allophones, the voiced bilabial stop ‘b’, which appears only after the homorganic nasal ‘m’, and the voiced bilabial fricative a, realized intervocalically, are also written the same way, using the voiced bilabial stop symbol ‘b’. Although the language has only one liquid, both ‘r’ and ‘l’ are used in the orthography. The liquid ‘r’ is used in all texts and ‘l’ is used only in loan words that have ‘l’ in their spelling, such as *Libiya* ‘Libya’, *Alijeriya* ‘Algeria’, *dolari* ‘dollar’.

Vowels and Consonants

Kinyarwanda has five vowels, which are either long or short and are high-toned or have no tones. The high tone can appear on either the first mora or the second mora. These vowels are the two high vowels ‘i’ and ‘u’, the midvowels ‘e’ and ‘o’, and the center low vowel ‘a’. The midvowels ‘e’ and ‘o’ are not allowed in both the (pre)prefix and suffix positions. In verbs, however, these midvowels can appear in the suffix position as a result of vowel harmony if the vowel of the verb stem is a midvowel, e.g., *gukosa* ‘to make mistakes’, *guko-soora* ‘to correct’ /ku-kos-uur-a/; *kumenya* ‘to know’, *kumenyeesha* /ku-meny-iish-a/.

The majority of word stems have the same identical vowel in all syllables: *u-mu-biri* ‘body’, *u-bu-riri* ‘bed’, *i-ki-reenge* ‘leg’, *i-béere* ‘breast’, *u-mu-góongo* ‘back’, *u-mu-hoondo* ‘yellow’, *u-ku-guru* ‘leg’, *u-ru-túgu* ‘shoulder’, *igihaánga* ‘skull’, *i-ki-gaanza* ‘hand palm’. This observation raises the question as to whether the stem is assigned only one vowel that is copied or that spreads to other syllables.

Since Kinyarwanda has open syllables only, loan words with cluster consonants copy the vowel of the syllable on the right or a default vowel ‘u’ with bilabial consonants and ‘i’ with other consonants. As the following loan word *porogaramu* ‘program’ shows, both vowels ‘o’ and ‘a’ are copied on the preceding vowelless consonants and the vowel ‘u’ is inserted after the final consonant ‘m’.

‘program’ /p.\$ro.\$g.\$ra.\$m.\$/ > *porogaramu*

This language has both simple and complex consonants. The simple consonants, using the official orthography, are the bilabials ‘p’, ‘b’, and ‘m’, the interdental ‘f’ and ‘v’, the alveolars ‘t’, ‘d’, ‘s’, ‘z’, and ‘n’; the alveopalatals ‘sh’, and ‘j’, and the velars ‘k’, ‘g’, and ‘h’. Kinyarwanda has two glides, the palatal ‘y’ and the bilabial ‘w’. It has one liquid, ‘r’, which is written as ‘l’ in some loan words, as was pointed out earlier. The affricates are the labiodental ‘pf’, the alveolar ‘ts’, and the palatal ‘c’.

The complex consonants are the prenasalized simple consonants, the palatalized consonants, the velarized consonants, the palatalized–velarized consonants, the prenasalized palatalized consonants, and the prenasalized palatalized–velarized consonants.

Prenasalized consonants are the bilabial ‘mp’ and ‘mb’; the interdental ‘mv’, ‘mf’, and ‘mpf’; the alveolar ‘nt’, ‘nd’, ‘ns’, ‘nz’, and ‘nts’; the palatal ‘nsh’, ‘nj’, and ‘nc’; and the velar ‘nk’, ‘ng’, and ‘nshy’.

Palatalized consonants are the bilabial ‘by’, ‘py’, and ‘my’; the interdental ‘fy’; the alveolar ‘ty’, ‘dy’, ‘sy’, and ‘nny’; and the velar ‘cy’, ‘jy’, and ‘shy’.

The velarized consonants are the bilabial ‘pw’, ‘bw’, and ‘mw’; the interdental ‘fw’ and ‘vw’; the alveolar ‘tw’, ‘dw’, ‘sw’, ‘zw’, ‘nw’, ‘rw’, and ‘tsw’; the palatal ‘shw’, ‘jw’, ‘cw’, and ‘yw’; and the velar ‘kw’, ‘gw’, and ‘hw’. Palatalized–velarized consonants are the bilabial ‘byw’, ‘pyw’, and ‘myw’; the alveolars ‘tyw’, ‘dyw’, ‘syw’; and the velar fricative ‘shyw’. Palatalized consonants, velarized consonants, and palatalized–velarized consonants can in turn be prenasalized as shown in the following examples: ‘mbyw’ (prenasalized palatalized–velarized voiced bilabial), ‘mvyw’ (prenasalized palatalized–velarized voiced interdental, ‘nshyw’ prenasalized palatalized–velarized voiceless fricative velar, and ‘njyw’ (prenasalized palatalized–velarized voiced stop velar). The complex consonants in Kinyarwanda are discussed at great length in Kimenyi (2002) and Bizimana *et al.* (1998). It is still an open debate in phonetics and phonology as to whether these complex consonants are one with multiple articulators or a sequence of independent segments.

Tonology

Role of Tones

Tones are lexical, morphological, and syntactical. Lexical tones differentiate words that look alike segmentally as shown in (1), morphological tones play the role that segmental morphemes are assigned in other languages as illustrated in (2), whereas syntactic tones are assigned depending on where the word bearing the tone occurs in the noun phrase, verb phrase or the sentence as shown in (3):

- (1) *inda* ‘stomach’ <> *indá* ‘louse’, *ino* ‘toe’ <> *inó* ‘here’
- (2) *basoma* ‘they read’ <> *basomá* ‘who read’ <> *básoma* ‘when they read’
- (3) *baraaza bagakóra* ‘they come and work’ <> *baraaza bagakora akazi* ‘they come and do the work’

In (2), the lack of tone shows the present tense, and the high tone on the second syllable shows that the verb is a relative clause, whereas the high tone on the first syllable of the verb stem shows that the verb is a temporal or conditional clause. In (3), the verb loses its high tone because it is followed by a complement.

Tone Rules

Tone rules in Kinyarwanda were thought to be complicated. However, when looked at very closely, they are very simple. There is only one lexical high tone per morpheme. Some morphemes are toneless. Noun tone patterns differ from verb tone rules. Any noun can have a lexical high tone on any syllable of the stem, except the augment and the prefix. A verb, however, even when it is polysyllabic and has multiple suffixes, can have a high tone only on the first mora of the first syllable or the first of the second syllable of the stem. Other syllables are extraprosodic. When high tones are found there, they are stray tones, since they do not participate in tone rules such as the Meeussen rule, Beat Movement, Iambic Reversal, etc. The first syllable verb high tone assignment is lexical, whereas the second syllable high tone assignment is grammatical. The prosodic domain of tone rules application of both nouns and verbs is the left-most phonological tone and the first mora of the stem first syllable. Tone rules apply from right to left, whereas in the majority of languages whose tone rules have been studied, they apply from left to right.

Nouns can obtain a secondary and a tertiary tone. A secondary high tone is assigned on the first mora of the noun stem if the lexical high tone is at least two or more mora away from the first noun mora.

isáandukú ‘box’ ← /i-saandukú/, *inkókorá*
‘elbow’ ← /i-n-kokorá/
aug-box aug-CL9-elbow
abasásamígozí ‘murderers’
/a-ba-sas-a+i-mi-gozí/
aug-sub.pr.-make bed-asp+aug-CL4-rope
inshóberamáhaánga ‘idiomatic expressions’
/i-n-shober-a+a-ma-haánga/
aug-CL9-disorient-asp+aug-CL6-foreign countries

A noun can thus have only a maximum of three phonetic high tones.

What makes verb tones seem complex is the assignment of the tense–aspect–modality morphemes. Some tenses or moods erase lexical tones, thus making the whole finite verb toneless, or assign tones to toneless verb stems, making both toneless verb stems and high-toned verb stems look the same. As shown below, the verb stem *-kin-* ‘play’ and *-kór-* ‘work/do’ are neutralized, becoming toneless or both bearing a high tone in some tenses.

ntibagikora ‘they do not work anymore’ <>
ntibagikina ‘they do not play anymore’
baracyáakóra ‘they still work’ <> *baracyáakina*
‘they still play’
bakoré ‘they should work’ <> *bakiné* ‘they should play’.

The metrical domain for verb tone rules is the first mora of the first object pronoun and the first mora of the verb stem for lexical tones and the first mora of the first object pronoun and the first mora of the second syllable of the verb stem. Kinyarwanda is one of the Bantu languages that can have multiple object pronouns.

baranáhabíbamákoreeshereza ← /ba-ra-na-ha-bi-ba-mu-kór-iish-ir-ir-y-a/
they-t-also-there-it-him/her-them-do-appl-caus-appl-appl-caus-asp
‘they also make them do it for him/her there’.

Phonology

Phonological Rules Affecting Vowels

Phonological rules affecting vowels are vowel deletion, vowel coalescence, gliding, vowel harmony, vowel shortening, and vowel lengthening.

When a word or morpheme that ends with a vowel is followed by another one that also starts with a vowel, the final vowel of the word or morpheme on the left is always deleted.

Vowel coalescence takes place within a word if there is a sequence of two morphemes ending with the central low vowel ‘a’ and starting with the high vowels ‘i’ and ‘u’, respectively: ‘a + i’ becomes ‘e’ and ‘a + u’ becomes ‘o’, e.g., *améenyo* /a-ma-iinyo/ ‘teeth’.

Gliding takes place within words or clitics if in a sequence of two vowels the first one is a high vowel ('i', 'u') or a round vowel ('u', 'o'), thus becoming 'y' for the front high vowel and 'w' for round vowels, e.g., *harimó amáazi* > *harim^w áamáazi* 'there is water in it'; *i-ki-úuma* > *ik^yúuma* 'knife'; *lu-bu-oónko* > *ub^woónko* 'brain'. Vowel harmony affects vowels in the suffix position. If the suffix vowel is high ('i', 'u') it becomes mid ('e', 'o', respectively), if the word stem vowel is a mid vowel, e.g., *gukóra* /ku-kór-a/ 'to work' > *gukóresha* /ku-kór-iish-a/ 'to cause to work/employ/use', *kumenya* /ku-meny-a/ 'to know' > *kumenyeesha* /ku-meny-iish-a/ 'to let know/inform'. Vowels in Kinyarwanda are always short in the beginning and final positions of words. They always lengthen before prenasalized consonants and after palatalized and labio-velarized consonants (Kimenyi, 1979, 2002).

Phonological Rules Affecting Consonants

Phonological rules affecting consonants are assimilation, dissimilation, known as Dahl's Law, fricative spread, deletion, and insertion. As in other Bantu languages, reduplication is also very productive in Kinyarwanda.

Consecutive consonants acquire the same voice, manner, and place of articulation phonetic features. Nasals take the place of articulation (labial, velar, palatal, velar) of the consonant on the right. Consonants obtained through the palatalization or velarization process also agree in voice, nasality, and place of articulation with the governor consonant.

gukúbita 'to hit' > *gukúbitkwa* 'to be hit' /ku-kubit-w-a/

kudóda 'to sew' > *kudódgwa* 'to be sewed' /ku-dód-w-a/

kubóna 'to see' > *kubónmwa* 'to be seen' /ku-bón-w-a/.

If a word has a palatalized fricative in one of the syllables on the right, fricatives in preceding syllables become palatalized as well.

gusoonza /ku-soonz-a/ 'to be hungry' >
gushoonjeesha /ku-soonz-iish-a/ 'to cause hunger'

basuuzugura /ba-suuzugur-a/ 'they despise' >
bashuujuguje /ba-suuzugur-ye/ 'they just caused to despise'.

This phenomenon argues for the autosegmental treatment of phonological rules because as the provided examples show, these fricatives do not have to be in adjacent syllables.

Reduplication is both lexical and grammatical. Lexical reduplication consists of stems that are already reduplicated. Grammatical reduplication affects the

stem. In verbs, it is very productive with verbs of movement or sound to show repetition, iterativity, or intensity.

Reduplication is achieved by either repeating the first syllable or the whole stem.

gutuumba 'to swell' → *gututuumba* 'to start swelling'

kugeenda 'to go'/'walk' → *kugeendageenda* 'to walk around'

ukwéezi 'moon/month' → *icyézezezi* 'moonlight'

ubusá 'nothing' → *ubusáabusá* 'very little quantity'.

Morphology

Noun Morphology

Kinyarwanda has 16 classes. Modifiers (adjectives, demonstratives, numerals, possessives) agree with the head noun by taking this class marker.

In some cases, however, the class marker has different phonetic forms depending on the grammatical category of the modifier, as illustrated in Table 1. The sentence in (2) in Table 1 shows how this type of noun class agreement works. The head noun *abagabo* (a-ba-gabo) 'men' with class 2 prefix-ba-has it copied to all modifying elements (adjectives, subject pronouns, object pronouns, etc.).

bá-no *bá-gabo* *bá-tatu* *bá-gufi*,
these men three short
mu-ra-bá-bon-a, *ba-mez-e néézá b-óose*.
you-pres-them-see-asp they-are-asp well all
'These three short men, you see them, they are all of them doing well'.

Table 1 Allomorphic variation of the nominal prefix according to its function

| | Noun | Adjective | Object pronoun | Demonstive | Possessive |
|-----|-------|-----------|----------------|------------|------------|
| 1. | u-mu- | mu- | -mu- | u- | u- |
| 2. | a-ba- | ba- | -ba- | ba- | ba- |
| 3. | u-mu- | mu- | -wu- | u- | u- |
| 4. | i-mi- | mi- | -yi- | i- | i- |
| 5. | i-ri- | ri- | -ri- | ri- | ri- |
| 6. | a-ma- | ma- | -ya- | a- | a- |
| 7. | i-ki- | ki- | -ki- | ki- | ki- |
| 8. | i-bi- | bi- | -bi- | bi- | bi- |
| 9. | i-n- | n- | -yi- | i- | i- |
| 10. | i-n- | n- | -zi- | zi- | zi- |
| 11. | u-ru- | ru- | -ru- | ru- | ru- |
| 12. | a-ka- | ka- | -ka- | ka- | ka- |
| 13. | u-tu- | tu- | -tu- | tu- | tu- |
| 14. | u-bu- | bu- | -bu- | bu- | bu- |
| 15. | u-ku- | ku- | -ku- | ku- | ku- |
| 16. | a-ha- | ha- | -ha- | ha- | ha- |

Note: The numbers 1–16 correspond to traditional conventional Bantu noun classification.

The Use of the Preprefix The preprefix or augment usually does not have any semantic function. Some Bantu languages such as Kiswahili do not have it. In Kinyarwanda, it is deleted, after demonstratives, in the vocative case and in onomastics (name creation). Within certain words, however, its absence marks definiteness and its presence indefiniteness.

mugaanga ‘the doctor’ > *umugaanga* ‘a doctor’
munywáanyi ‘the buddy’ > *umunywáanyi* ‘a buddy’
mugeenzi ‘the friend’ > *mugeenzi* ‘a friend’
mwaarimú ‘the teacher’ > *umwáarimú* ‘a teacher’

The absence of the preprefix bleeds tone rules and its absence feeds them. The secondary tone assignment on the first more of the noun stem takes place only if it has a preprefix.

Noun Derivation with the Prefix Kinyarwanda nouns have a small number of suffixes. The most productive one is *-kazi*, which is added to the stem to show feminine, e.g., *umunyarwaanda* ‘Rwandan’ <> *umunyarwaandakazi* ‘female Rwandan’, *umwáarimú* ‘teacher’ <> *umwáarimúkazi* ‘female teacher’.

Derivation is productive with the preprefix that creates new words that are either metaphorically or metonymically related to the original noun as shown by the stem-*ntu* in the following examples: *umuuntu* ‘person’, *ikiintu* ‘object’, *ukuuntu* ‘manner’, *ubuuntu* ‘generosity’, *ahaantu* ‘place’.

Verb Morphology

The simple Kinyarwanda verb form consists of the subject pronoun, the verb stem, and the aspect marker. The aspect marker is either *-a(ga)* (imperfective aspect) or *-ye* (the perfective aspect as seen below. The *-aga* suffix is used in past tenses only and is not used in Kirundi.

basoma (ba-som-a) ‘they read’;
basomaga /ba-a-som-aga/ ‘they were reading’;
basomye (ba-som-ye) ‘they just read’.

The complex form consists of the preprefix, the subject pronoun, the tense–aspect–modality (TAM) morphemes, the object markers, the reflexive pronoun *-i-*, the verb stem, the lexical verb extensions, the grammatical morphemes, the aspect marker and the postsuffixes *-mó*, *-hó*, or *-yó*.

The preprefixes are either the morpheme *nti-* or *ni-*, negative and temporal morphemes, respectively. The TAM morphemes show time, mood or aspect. Two TAM morphemes can occur in the same slot. Kinyarwanda can have multiple object pronouns, multiple lexical verbal extensions, and multiple grammatical suffixes. Lexical extensions such as *-agur-*, *-iir-*, *uur-*,

-aang, *iriz-*, etc., add lexical information, such as inchoativity, iterativity, repetitivity, intensity, frequentativity, reversivity. Grammatical morphemes, such as the causative morpheme *-iish-*, the applicative morpheme *-ir-*, the comitative/reciprocal morpheme *-an-*, can be added to any verb stem. The following sentence serves as an example to illustrate a verb with multiple object pronouns and multiple grammatical suffixes:

Umugoré a-ra-na1-ha2-ki3-zi4-ba5-ku6-n7- som-eesh-eesh-er-er-eza

woman she-pres-also1-there2-it3-it4-them5-you6-me7-read-caus-caus-appl-appl-aspl

‘The woman is also making them use it to do it for me for me there’.

Lack of Adjectives Kinyarwanda has a handful of adjectives (less than 20). What is expressed by adjectives in other languages is rendered by either the possessive construction (X of Y) or the relative clause construction.

| | | |
|-------------------|----------------|------------------|
| ‘a poor person’ : | <i>umuuntu</i> | <i>w’úmukené</i> |
| | person | of poor |
| | <i>umuuntu</i> | <i>ukénnye</i> |
| | person | who-is poor |

Ideophones Ideophones are common not only in Bantu languages but also in the whole Niger–Congo language family. They are different from onomatopoeias, which imitate sounds of nature. They can express different concepts that do not have anything to do with sound by using sound symbolism, short or long vowels, reduplication, triplication, or quadruplication. They can also have different grammatical functions.

umuseké *weerá de*.
dawn which-is clear (ideophone)
‘a very clear dawn’
icyáayi *tsiritsiri*
tea ideophone
‘a very dark coffee’.

Unclassified Categories As was pointed out earlier, Kinyarwanda has a handful of adjectives. It is the same with function words as well, namely, auxiliaries, prepositions, conjunctions, and subordinators. These are expressed by noun phrases or verb phrases. In most cases, the structure tells whether the noun or verb is the noun or the verb or a function word.

muu nsii *y’(á)am(é)ez(á)*
under earth of table
‘under the table’

baravúgana usíbye kó batabonána.

they-talk-to-each-other you-are-absent that they-do
not see each other
'they talk to each other except that they do not see
each other'.

Stems of nouns, verbs, and unclassified words can have different phonetic variations, as the word for diploma shows:

diploma: *dipóroómi, dipóromá, dipóromé, dipóromí, dipóromó, dipóromú, dipóroóma, dipóroóme, dipóroómu, dipóroómo, diipóroómu, diipóroómo, diipóroóme, diipóroómi, diipóroóma.*

It is still an unsolved question for Kinyarwanda lexicographers to decide which form should be considered the main form or whether all forms should entered in the dictionary as independent lexical entries.

Syntax

Kinyarwanda, like other Bantu languages, is a SVO language. Modifiers follow head nouns. What is interesting about this language, as pointed out in Kimenyi (1980, 2002), is the existence of (a) the subject–object reversal, (b) the wh-question *in situ*, (c) the lack of relative pronouns, (d) serialization, and (e) the existence of multiple direct objects.

Object–Subject Reversal and Existential Construction

The object–subject reversal consists of interchanging the object and the subject positions, whereas the existential construction puts both the subject and the object after the verb, prefixing the verb with the locative morpheme *ha-*(CL16). Neither construction changes the meaning, except that focus is on the object.

Umwáana a-ra-som-a igitabo.

'The child is reading the book'.
child s/he-t-read-asp book

Igitabo ki-ra-som-a umwáana.

'The book is reading the child'.
book CL7-t-read-asp child

Ha-ra-som-a igitabo umwáana.

'It is the child who is reading the book'.
CL16-t-read-asp book child

The object–subject reversal and the existential constructions have the same function as the passive, which is shown by the suffix *-w-* added to the verb just before the aspect marker.

Igitabo ki-ra-som-w-a n'úumwáana.

'The book is being read by the child'.
book CL7-t-read-pass-asp by child

Wh-Question *In Situ*

In Kinyarwanda and many other Bantu languages, wh-questioning is only allowed *in situ*.

W-iit-w-a nde?

you-call-pass-asp who
'you are called who?' > 'What is your name?'

Ba-tuu-ye he?

they-live-asp where
'Where do they live?'

Lack of Relative Pronouns

Kinyarwanda does not have relative pronouns. Relative constructions are marked by a high tone on the verb stem instead.

Abáana ba-som-á ibitabo.

children they-read-asp/rel books
'The children who read books'.

Ibitabo abáana ba-som-á.

books children they-read-asp/rel
'The books that the children read'.

Serial Verb Construction

When multiple verbs precede the sentence main verb, they lose their semantic function and serve as auxiliaries or tense–aspect–modality bearers. This is illustrated by the following sentences:

| | | |
|------------------------------------|-------------------|---------------------|
| <i>Ba-a-ri</i> | <i>bá-tuu-ye</i> | <i>bá-saanz-w-e</i> |
| they-t-be | they-dwell-asp | they-join-pass-asp |
| aux ¹ | aux ² | aux ³ |
| <i>bá-jy-a</i> | <i>bá-kuund-a</i> | <i>gu-pf-a</i> |
| they-go-asp | they-like-asp | to-die |
| aux ⁴ | aux ⁵ | aux ⁶ |
| <i>ku-dú hamagar-a.</i> | | |
| to-us-call-asp | | |
| V | | |
| 'They usually at least called us'. | | |

| | | |
|--------------------------------------|-------------------|-------------------|
| <i>Mu-siga-ye</i> | <i>mú-geend-a</i> | <i>mú-beerako</i> |
| you-stay-asp | you-walk-asp | you-start |
| aux ¹ | aux ² | aux ³ |
| <i>mú-du-subiz-a.</i> | | |
| from you-us-answer-asp | | |
| V | | |
| 'Now you respond to us immediately'. | | |

Multiple Direct Objects

Kinyarwanda, like many other Bantu languages, can have multiple direct objects. These objects are either inherent or structural.

Recipients or benefactives are introduced directly to the verb without any preposition with some interactive verbs (giving, showing, etc.).

Umugabo a-haa-ye abáana ibiryó
man he-give-asp children food
'The man has just given food to the children'.

Umuwáarimú a-r-éerek-a abanyéeshuúri amashusho.
teacher he-t-show-asp students pictures
'The teacher is showing pictures to the students'.

Inalienable possessions can appear as direct objects without any verb extension marker.

Umugoré a-ra-kúbit-a umwáana ukuguru n'inkoni.
woman sub.pr.-t-hit-asp child leg with stick
Umugoré a-ra-kúbit-a umwáana inkoni ku kuguru.
woman sub.pr.-t-hit-asp child stick on leg
'The woman hitting the child on the leg with a stick'.

As shown by these examples, the inalienable possession 'ukuguru' and the instrumental 'inkoni' can appear as either adjuncts or direct objects without any verbal extension.

Structural direct objects are obtained by deleting prepositions of adjunct objects and by adding suffixes such as -iish-, -ir-, or -an- to the verb stem.

Umugabo a-ra-andik-a ibáruwá n'íkáramú
man sub.pr.-t-write-asp letter with pen
Umugabo a-ra-andik-iish-a ikáramú ibáruwá
man sub.pr.-t-write-caus-asp pen letter
'The man is writing a letter with a pen'.

Conclusion

Kinyarwanda is a prototypical Bantu language. It has all the features that characterize this language group. Its main contributions in syntax have been about the nature and function of grammatical relations (Kimenyi, 1980) and in tonology, its contributions have been about the nature of tone, tone representations, and tone rule application (Kimenyi, 2002).

Bibliography

- Bizimana S *et al.* (1998). *Imiteérere y'Ikinyarwanda*. Kigali, Rwanda: Pelloti Set Press.
- Coupez A (1980). *Abrégé de Grammaire rwanda. Tome I and Tome II*. Butare, Rwanda: Institut National de Recherch Scientifique.
- Kimenyi A (1979). *Studies in Kinyarwanda and Bantu Phonology*. Edmonton, Alberta: Linguistic Research Inc.
- Kimenyi A (1980). *A relational grammar of Kinyarwanda*. Berkeley/Los Angeles: University of California Press.
- Kimenyi A (2002). *A tonal grammar of Kinyarwanda: An autosegmental and metrical analysis*. Lewiston, NY: The Edwin Mellen Press.

Kirghiz

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Kirghiz (*qıryız tili*, *qıryızça*) belongs to the North-western or Kipchak branch of the Turkic language family, more specifically, to its Southern or Aralo-Caspian group. Until the early 20th century, it was called Kara-Kirghiz, whereas Kazakh was referred to as Kirghiz or Kazak-Kirghiz. Kirghiz is spoken in the Kyrgyz Republic (*Qıryız Respublikası*) or Kyrgyzstan (*Qıryızstan*) and in parts of Uzbekistan, Tajikistan, China (Xinjiang), the Russian Federation, Kazakhstan, etc. Its main area is the mountainous part of Western Turkistan, the plateaus of the western Tien-shan south of Kazakhstan, and the Alay mountain south of Ferghana. The number of speakers amounts to about 3 million, in Kyrgyzstan over 2.5 million.

In spite of the existence of a modern Kirghiz standard language, Russian has remained the dominant

language of higher education, administration, and so forth in the Republic. Since Kirghiz was proclaimed the official language of Kyrgyzstan in 1989, it has consolidated its position, acquiring more social functions. In 1996, Russian was made an official language, along with Kirghiz, in territories and workplaces in which Russian-speaking citizens predominate.

Origin and History

It is still unclear to what extent the Kirghiz of today are successors of the Old Kirghiz, the first Turkic people mentioned in Chinese sources and described there as blond and blue-eyed. This group settled on Upper Yenisey. Runiform inscriptions found on the territory of today's Tuva indicate that the first Kirghiz state was, at the beginning of the 8th century A.D., located north of the Sayan mountains. In 840 the Kirghiz ended the old steppe Uyghur empire and established their own empire, which lasted until 920. Most old Turkic groups left this region at the turn of the millennium. A few groups remained in Siberia, (e.g., the ancestors of the Kirghiz and the Altay Turks).

Some Kirghiz tribes may already have moved to the Tianshan region by the 10th century. Other tribes followed, particularly during the Mongol attacks. The Mongol expansion in the 13th century forced old Kirghiz groups to migrate to Western Turkistan and the Tianshan region.

In the following centuries, the tribes referred to as Kirghiz were gradually pushed back by Oirats and Dzungars. In the 16th century, the Kirghiz acted as an important ally of the Kazakh. In the second half of the 17th century, the Yenisey Kirghiz were forced to accept the sovereignty of the Kalmyk. At the beginning of the 18th century, the majority of the Kirghiz migrated to Tianshan. After the breakdown of the Dzungar Empire in 1758, the Kirghiz definitively settled in their present territory. In the 18th and 19th centuries, they were subjects of the Uzbek Khanate of Kokand. They came under Russian supremacy in 1867, and their territory was incorporated into the Russian Empire in 1880. Under the Russian and Soviet rule, numerous Kirghiz emigrated to China. In 1991, the Kirghiz Republic was proclaimed an independent state.

Related Languages and Language Contacts

Kirghiz is closely related to Southern Altay Turkic of South Siberia. Its modern form is also very close to Kazakh as a result of long-standing intensive contacts. Old Kirghiz, as attested to in inscriptions, was similar to Orkhon Turkic and Old Uyghur (see *Turkic Languages*). The language was later influenced by Mongolic and especially by Kipchak Turkic. In the 18th and 19th centuries, it was subject to some impact from the Iranicized dialects of the Uzbek area. The contacts with Russian began at the end of the 19th century. After the Kirghiz territory was conquered by the Russian empire in the second half of the 19th century, the Russian influence became predominant.

The Written Language

In the Soviet period, a Kirghiz standard language was developed on the basis of the northern dialects. Before the revolution Kirghiz had already found some limited use as a written language. A modified version of the Arabic script was introduced in 1924 but was given up in 1928 in favor of the unified Roman-based alphabet. A modified Cyrillic-based script was introduced in 1940. In the post-Soviet era, a new Roman-based alphabet was created, but it has not yet replaced the Cyrillic-based script. The Arabic script is used for the variety written in China.

Distinctive Features

Kirghiz exhibits most linguistic features typical of the Turkic family (see *Turkic Languages*). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation; for example, **A** = **a/e**, and **I** = **i/i**. Hyphens are used here to indicate morpheme boundaries.

Phonological Features

Kirghiz has a rather regular vowel system, lacking the reduced vowels of Kazakh. The different orthographic vowel representations very often conceal the phonetic similarities between Kirghiz and Kazakh.

Kirghiz exhibits the typical Kipchak labialization of **g/ɣ**, but does not, as Kazakh, replace them by the glide **w**. As in Altay Turkic, the preceding vowel is deleted and lengthened, as in **to:lu**: [mountain-DER] 'mountainous' (< **ta:ɣ-l°ɣ**), cf. Altay Turkic **tu:lu**: [mountain-DER], Kazakh **taw-lı** [mountain-DER]. Kirghiz differs from Kazakh by absence of the sound changes **č > š** and **š > s**. Modern Kirghiz displays initial **j̥**- instead of older **y**- (e.g., **jol** 'way'; cf. Turkish **yol**). This change occurred under the influence of Kazakh, which, however, later changed **j̥** to **ž**. Kirghiz does not display the uvular **x** or the glottal **h** but replaces them by **q** or zero in loanwords (e.g., **qabar** 'message' [< **xaber**], **ar** 'each' [< **har**]). Kirghiz has, like some Siberian Turkic languages, a well-developed sound harmony. Suffixes exhibit both front vs. back harmony and rounded vs. unrounded harmony. The choice of suffix vowels is determined by features of the preceding syllable (e.g., **köl-dör-dön** [lake-PL-ABL] 'from the lakes,' **üy-lör-übüz-dö** [house-PL-POSS.1.PL-LOC] 'in our houses'). The rounded vs. unrounded harmony also affects low suffix vowels, which are rounded to **o** and **ö** after **o**, **ö**, **ü** in the preceding syllable. However, **u** in the preceding syllable is not followed by **o** (e.g., **qum-da** [sand-LOC] 'in the sand'). Rounding of low suffix vowels is also observed in Altay Turkic, Turkmen, Bashkir, and to some extent, in Kazakh, and Noghay.

Morphophonemic alternations are found in suffixes with initial **l** and **n**, which are assimilated to **d** after voiced consonants (with some exceptions after **r** and **y**), and to **t** after voiceless consonants (e.g., **ata-nın** [father-GEN] 'of the father,' **qar-dın** [snow-GEN] 'of the snow,' **at-tın** [horse-GEN] 'of the horse,' **alma-lar** [apple-PL] 'apples,' **kün-dör** [day-PL] 'days,' **at-tar** [horse-PL] 'horses').

Arabic and Persian loanwords have generally been adapted more strongly to the native phonological

system than their counterparts in neighboring Turkic languages (e.g., **apta** ‘week’ [< **hafta**], **ubaqtı** ‘time’ [< **waqt**]).

Grammar

The Kirghiz genitive suffix **-nIn** ends in **-n**, as in languages of the Southwestern (Oghuz) branch, not in **-ŋ** as in the neighboring languages.

The suffix **-rA:K** (after consonant-final stems **-IrA:K**) is added to adjectives to express comparative degree. The superlative is expressed by means of the preposed element **eŋ** (e.g., **eŋ jaqşı** [SUPERL good] ‘best’). The normal second-person plural form of the personal pronoun is **si-ler** [you-PL] ‘you.’ **Siz** is the polite form used for one addressee, and **siz-der** [you-PL] is the polite form for more than one addressee. The corresponding copula suffixes are **-slŋAr**, **-slz** and **-slz-dAr**. Most demonstrative pronouns exhibit an optional **-l** in the nominative; for example, **bu(l)** ‘this.’ Different degrees of closeness are marked with **bu(l)** (close referents known to the speaker), **oŋo(l)** and **uŋul** (more remote referents), and **tigi(l)** and **tetigil(l)** (remote referents outside the conversational setting).

Pronouns also include the reflexive **öz** and the interrogatives **emne** ‘what,’ **kim** ‘who,’ **qaysı** ‘which,’ and so on. A collective suffix **-O:** (with drop of stem-final vowels) is used with the numerals from 1 to 7 (e.g., **ek-ö:** [two-COLL] ‘two together’). Distributives are formed with the ablative (e.g., **otuz-dan** [thirty-ABL] ‘thirty each’). Approximative and multiplicative numbers can be expressed with **-dAy**, **-čA** and **-LA-GAn** (e.g., **otuz-day** [thirty-APPR], **otuz-ča** [thirty-APPR] ‘about thirty,’ **jüz-dö-gön** [hundred-MULT] ‘hundreds’).

As in most other Turkic languages, the third-person possessive suffix exhibits the so-called ‘pronominal **n**’ (e.g., dative **-n-A**, locative **-n-dA**). The past copula particle is **ele** ‘was,’ instead of **edi** ~ **idi** in other Turkic languages.

The cooperative-reciprocal form in **-(I)š** is also used to indicate the third-person plural of finite verbs (e.g., **jaz-iš-at** [write-PL-PAST.3] ‘they write’ [singular **jaz-at**]), **kel-iš-ti** [come-PL-PAST.3] ‘they came’ [singular **kel-di** [come-PAST.3.SG]]. A general present tense is formed with **-A** + personal markers (e.g., **bar-a-t** [go-PRES-3.SG] ‘goes’). A more focal present tense (with a narrower focus on the ongoing event) is formed with **-convert** + present forms of **jat-** ‘to lie’ (and three other auxiliary verbs) + personal markers (e.g., **bar-a jat-a-t** [go-CONV AUX-PRES-3.SG] ‘is going’). The suffix **-Dlr** adds a presumptive meaning (e.g., **oyyon-yon-dur** [wake up-POSTTERMINAL-PRES.3.SG] ‘has presumably waken up’). An evidential past is formed with **-(I)p-tlr**. A habitual or

durative past is formed with **-čl** + personal markers (e.g., **kel-čü-büz** [come-HABIT.PAST-1.PL] ‘we used to come’). Intention is expressed by **-MAK-čl** (e.g., **kel-mek-či-min** [come-INTENT-1.SG] ‘I want to come’). The system of conjunctions is weakly developed, as Kirghiz has not been under strong Iranian influence.

Lexicon

The basis of the Kirghiz vocabulary consists of Kipchak Turkic elements. Similar to the languages of other nomadic groups, Kirghiz has borrowed numerous Mongolic words as a result of close contacts, especially in the Middle Ages (e.g., **dülöy**, ‘deaf’; **belen**, ‘ready’; **qara-**, ‘to look’). Arabic and Persian words, copied via Chaghatay and Uzbek particularly into the southern dialects, constitute a sizable part of the vocabulary, covering various domains of Islamic culture (e.g., **künö**, ‘sin’; **šar:**, ‘city’; **baqča**, ‘garden’; **pikir**, ‘thought’). The northern dialects, on which the literary language is based, are less influenced by the Islamic vocabulary.

Russian loanwords, which constitute the most recent layer in the lexicon, were introduced from the end of the 19th century on. The use of Russian words is rather dominant in informal spoken standard Kirghiz. Efforts have been made in the last decades to reduce the amount of Russian terms by creating neologisms on the basis of Turkic and Arabic–Persian lexical material or on Russian models. Neologistic suffixes include **-čIl** (e.g., **ulut-čul** [nation-DER] ‘nationalist’; from **ulut**, ‘nation’), **-Glč** (e.g., **uč-quč** [fly-DER] ‘pilot’ [from **uč-** ‘to fly’]).

The variety of Kirghiz spoken in Xinjiang exhibits numerous Chinese loanwords. The written language, however, is largely oriented toward the norm used in Kirghizstan.

Dialects

The Kirghiz dialects can be divided into a southern and a northern group, the latter forming the base of the standard language. Northern dialects tend toward intervocalic voicing of **s** (e.g., **bala-zı** [child-POSS.3.SG] ‘his/her child’ instead of **bala-sı** [child-POSS.3.SG]), a feature typical of the South Siberian Turkic languages. The southern dialects, mainly those spoken in the Ferghana basin, show different degrees of Uzbek influence. They lack some characteristic features that separate Kirghiz from other Turkic languages. For example, the glide **w** is preserved in cases in which the standard language only exhibits a long vowel as a trace of a velar that once underwent labialization (e.g., **tow** ‘mountain’

instead of *to*.) The past copula particle *ede* 'was' is used instead of *ele*. The plural suffix is used in third-person plural forms of finite verb (e.g., *kel-di-ler* [come-PAST-PL.3] 'they came'), instead of *kel-iš-ti* [come-PL-PAST.3]. The 'pronominal *n*' has been lost under Southeastern Turkic influence. There are also other changes in the nominal inflection. Words copied from Arabic and Persian via Chaghatay and Uzbek have preserved their original phonetic shape to a higher degree than in the northern dialects.

Bibliography

Hebert R J & Poppe N (1963). *Kirghiz manual*. The Hague: Mouton.

Imart G (1981). *Le kirghiz: 1-2*. Aix-en-Provence: Publications de l'Université de Provence.

Johanson L (2001). 'Kirghiz.' In Garry J & Rubino C (eds.) *Facts about the world's major languages: an encyclopedia of the world's major languages, past and present*. New York: H. W. Wilson, New England Publishing Associates. 387–389.

Judaxin K K (1965). *Kirgizsko-russkij slovar'*. Moskva: Sovetskaja Ėnciklopedija.

Kirchner M (1998). 'Kirghiz.' In Johanson L & Csátó É Á (eds.) *The Turkic languages*. London: Routledge. 344–356.

Wurm S (1949). 'The (Kara-)Kirghiz language.' *Bulletin of the School of Oriental Studies* 13, 97–120.

Zaxarova O V (ed.) (1987). *Grammatika kirgizskogo literaturnogo jazyka 1: Fonetika i morfologija*. Frunze: Ilim.

Kordofanian Languages

T C Schadeberg

© 2006 Elsevier Ltd. All rights reserved.

Kordofanian is the name of an African language family. It derives its name from that of a former Islamic state with El Obeid as its center.

All 20-odd Kordofanian languages are spoken in the Nuba Mountains in the Republic of the Sudan. The total number of speakers is estimated at around 200 000. Available information about Kordofanian languages is sketchy; no Kordofanian language has been well documented and analyzed. The unity of the Kordofanian language family was first postulated by J. H. Greenberg in 1950; later (in 1963), he classified Kordofanian as a primary branch of his Niger–Congo family.

There are four branches of Kordofanian, named after centrally located towns (names of individual languages are given in parentheses): Heiban (Moro, Tiro, Shirumba, Utoro, Ebang, Laru, Logol, Rere, Warnang, Ko), Talodi (Ngile, Dengebu, Tocho, Jomang, Nding, Tegem), Rashad (Tagoy and Tegali dialect clusters), and Katla (Kalak, Lomorik). Data from wordlists and short grammatical descriptions make it clear that at least the first three branches of Kordofanian have a noun class system (marked by prefixes), which may be taken as evidence for genetic relationship (i.e., common origin) with the large

Niger–Congo language family. Lexical evidence for this relationship remained very limited in the early 1990s.

The nine languages of the Kadugli group (Yega, Mudo, Talla, Miri, Tolubi, Kufo, Sangali, Krongo, Talassa) are spoken by about 100 000 people living on the hills lining the southern edge of the Nuba Mountains. One of these languages has been described in a monograph (Reh, 1985). Greenberg originally classified the Kadugli languages as being part of Kordofanian, but this view has since been challenged. The Kadugli languages have systems of nominal classification that distinguish three or four genders, but any existing typological or substantial similarities with Niger–Congo are not sufficient to claim genetic relationship. It seems most likely that Kadugli belongs to the Nilo–Saharan language family – just as do all the other languages that surround Kordofanian.

Bibliography

Greenberg J H (1963). *The languages of Africa*. The Hague: Mouton.

Reh M (1985). *Die Krongo-Sprache (niino mo-di): Beschreibung, Texte, Wörterverzeichnis*. Berlin: Dietrich Reime.

Schadeberg T C (1989). 'Kordofanian.' In Bendor-Samuel J (ed.) *The Niger–Congo languages*. Lanham, MD: University Press of America.

Korean

J H Grayson, University of Sheffield,
Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

Phonology

Basic Structures

The Korean language is a nontonal, polysyllabic, agglutinative language belonging to the Altaic family and probably closely related to the Manchu and Tungus members of that language family. The only major modern language to which Korean would appear to be related is Japanese, but the two languages, although similar in most respects grammatically, are significantly different phonologically. Korean and Japanese are therefore linguistic isolates due to the lack of sources to demonstrate the precise linguistic connections between them and with other members of the Altaic family.

Unlike Chinese, Korean lacks true tonal sounds, although it does have vowel stress. The morphological structures of Korean are extremely complex. Korean vocabulary items are built up of multiple morphemes into a highly polysyllabic composition. Like all members of the Altaic family of languages, Korean uses certain morphemes as functional markers to indicate the role of a word within the sentence, as well as mood, tense, location, and the social relationship between the speaker, listener, and the person spoken about.

Triple Consonantal Structure

The consonants of the Korean language are unusual for the triple distinction that is made between soft consonants (lenis consonants), hard, unaspirated consonants, and hard, aspirated consonants. The consonants of the lenis series are *k*, *n*, *t*, *l*, *m*, *p*, *s*, and *ch*. The hard, unaspirated consonants are *kk*, *tt*, *pp*, *ss*, and *tch*. The hard, aspirated consonants are *k'*, *t'*, *p'*, *ch'*, and *h*. (These transcriptions follow the orthographic conventions of the McCune-Reischauer System of Romanization, the standard system of scholarly transcription.) The sound of *l* becomes a strongly flapped *r* when placed in an intervowel context. Usually consonants of any of the three series of consonants are pronounced as voiceless, with the exception that the soft consonants *k*, *t*, *p*, and *ch* are pronounced *g*, *d*, *b*, and *j* when they occur between voiced sounds.

Consonantal Position

A principal phonological feature of Korean is the extreme restriction of consonant position within a given morpheme. Certain sound sequences within a morpheme are not permitted, such as *s* combined with *k*, although the reverse may occur when *k* is final in the preceding morpheme and *s* begins the succeeding morpheme. Certain consonants when in the final position in the morpheme become a strongly dentalized sound. Thus, *t*, *tt* fortis, unaspirated *t*, *t'* (fortis, aspirated *t*), *s*, *ch*, and *ch'* (fortis aspirated *ch*) all are pronounced as if they were *t* when they occur in the final position.

Intermorphemic Sound Change

Sound change between syllables is an important feature of the pronunciation of Korean morphemes. This feature, also true of Japanese, is made more difficult for the reader of Korean because it is an orthographic convention that the shape of the individual syllable (morpheme) should be preserved. Although the Korean alphabet itself is highly phonetic, the orthographic convention to preserve the written appearance of the syllable means that the reader must learn a large number of standardized sound changes that occur in the intersyllable position.

Intermorphemic Sound Movement

Sound movement between syllables also occurs. When a syllable that ends in a consonant is followed by a syllable beginning with a vowel, the final consonantal sound passes over to the next syllable. This passage of sound is not represented orthographically.

Nonclustering of Initial Consonants

Clusters of consonants at the beginning of a syllable are not characteristic, there being no equivalents of English *sk*, *st*, *str*, *sh*, and so on.

Triple Vowel System

The vowel system of Korean is as complex as the system of consonants. There are three ranges of vowels: standard vowels (monophthongs), vowel sounds beginning with *y* (rising diphthongs), and a wide range of full diphthongs and diphthongs beginning with the sound *w*. The basic vowels of the monophthong series are pronounced similarly to the vowels of the Romance languages. The monophthongs are *a*, *ö*, *o*, *u*, and *ü*. The *y* series of

rising diphthongs are *ya*, *yŏ*, *yo*, *yu*. The principal diphthongs number 11, although other combinations are possible. Vowels are characterized by phonemic length, which refers to an alteration in tonal height. There is evidence of an earlier stage of vowel harmony that exists as a residual characteristic in certain linguistic contexts.

Grammar and Syntax

General Features

Korean is an agglutinative language with strong elements of fusion and analytical development. The morphological development of word derivation is a well-developed feature of the grammar of the language. Nouns possess a wealth of case forms, possess the grammatical category of specification, and do not possess grammatical gender. There are forms of demonstrative pronouns that indicate varying degrees of spatial relationship.

Number

There are two types of numeral systems: the indigenous Korean system, and the Sino-Korean system that was borrowed as an entire loanword system. Along with the numeral system, there is a system of classifiers that are bound morphemes used as counting words to refer to objects, animals, or people.

Syntactical Markers

The predicatives, verbs, and adjectives of Korean do not have person, number, or gender. They do possess markers indicating social status, tense, and a sentence conclusion. There are three major bands of social status or reference that can be indicated with the special markers, each band containing within it possibilities for further refinements to indicate the precise degree of social relationship existing between the speaker, the listener, or the person spoken about. Tense markers indicate three broad classes of time: the present, the past, and the future (more properly, supposition about the occurrence of an event). Sentence conclusion markers indicate a wide range of moods and meaning, including simple declaration, interrogation, request, demand, suggestion, and reflection. In addition, there are quotative constructions that may be added to the verb to indicate the quotation of a declaration, interrogation, demand, or request. The structure of the verb is verb stem + honorific infix + tense infix + sentence conclusion marker (vs + hi + ti + scm). There is a separate lexical form of

the verb that is used to place the verb in alphabetical order in dictionaries, lexicons, and word lists.

Word Order and Word Relation

Word order in sentences for both independent and dependent clauses is always in the sequence of subject, object, predicate. Modifiers, whether of the adjectival or adverbial type, are always in the preposition modifying the word to which they refer. Syntactic relations between words may be expressed by postpositional markers, particles, syntactic nouns, adverbial particles, participles, and the infinitive form of predicatives. Thus, a sentence may consist of a series of clauses, such as an extended adjectival clause modifying a noun, which contains its own subject, object, and predicative with tense and honorific markers attached.

Speech Levels and Honorifics

As a key characteristic of the use of the Korean language is an appropriate use of the system of honorifics that show deference, a sentence must take into consideration three dimensions of speech relationship: (a) the nature of the relationship between the speaker and the listener, (b) the nature of the relationship between the speaker and the person spoken about, and (c) the appropriate way to speak of or about oneself. Any complete sentence will take into account at least one of these dimensions in addition to considerations of tense and mood.

Pronouns, especially for the second person, are very seldom used, the subject of the sentence being understood from the linguistic context.

Sentence Linkage

In sentences containing two independent clauses, the two clauses are linked together through a connection marker attached to the predicative of the first clause. The predicate of the final independent clause will contain markers for honorifics, tense, mood, and sentence conclusion. Where the initial clause is dependent, the connection marker attached to the predicative will indicate the precise relationship of the dependent clause to the independent and principal clause.

Vocabulary

Lexicon

Korean vocabulary is of three types: indigenous Korean vocabulary, Sinitic vocabulary, and loanwords

from European languages. Indigenous Korean vocabulary is highly polysyllabic in structure, a feature that is put to good use in sound imitation. Of the world's languages, Korean is one of the most highly onomatopoeic languages. Sinitic vocabulary consists of three subtypes: (a) uniquely Korean terms created by using Chinese characters, (b) direct loanwords from Chinese, and (c) Sino-Japanese loanwords. Sinitic vocabulary consists of terms that are in both ordinary and learned usage, and constitutes more than half of the entire Korean lexicon. European languages, particularly English, have contributed a number of words, both to the speech of the ordinary person and to the technical speech of professional persons. French, German, Spanish, and Portuguese have also made small contributions to the vocabulary of Korean. There are, or were, a small number of pure Japanese loanwords, but most of these have fallen into disuse through a movement for the purification of the language.

Word Creation

It is still common to use Chinese characters to create new items of technical vocabulary, for example, *nokhwa-gi*, for 'videotape recorder.' Most Sinitic items of vocabulary enter the language as nouns. By attaching the verb *hada* 'to do' in the postposition of the noun, loanwords of this type may be transformed into verbs. By using one of several constructions, such verbs may then be made into adjective or adverbial constructions. By adding *ki* (carrying a sense of continuous action) or *um/m* (carrying an abstract sense) to a verb stem, a verb may be nominalized. Dictionary definitions are often given using *um/m* attached to the final verb.

Parallel Vocabulary Sets

Throughout the vocabulary of Korean, there exists a parallel set of Korean and Sino-Korean vocabulary. Mention was made earlier of the existence of two systems of counting. This feature carries throughout the entire Korean lexicon. Often, but not exclusively, Sino-Korean words are used to name objects or subjects of discourse, while Korean words have a descriptive function. On some occasions, there is no preference in the use of one or the other type of vocabulary; in other instances, it is a matter of honorific or nonhonorific usage. With regard to time, hours

are given in Korean numbers, while minutes are given in Sino-Korean numbers. Again, duration of time (i.e., 'it took 1 hour to go home') is given using Korean numerals.

Notwithstanding the enormous impact that Sinitic vocabulary has had on enriching the vocabulary of the Korean language, there has been virtually no influence on the grammar of Korean, possibly because Chinese and Korean derive from two radically different language families.

The Written Language

Modern Korean may be written in one of two forms, either by using only the Korean alphabet (known as *Han'gul* in the Republic of Korea) or a mixed script of *Han'gul* and Chinese characters. In North Korea, the Korean alphabet is used exclusively. In South Korea, usage varies from context to context. A personal letter may use only the indigenous alphabet, while newspapers, textbooks, and better-quality popular books will use a large number of Chinese characters in the text. The more sophisticated and formal a piece of writing is, the more Chinese characters will be used. The Ministry of Education requires a high school graduate to have mastered between 1700 and 1800 characters, the number of characters expected to be encountered in a daily newspaper.

Bibliography

- Baek E-J (1984). *Modern Korean syntax*. Seoul: Jung Min.
 Dupont R & Millot J (1984). *Grammaire coréenne*. Paris: Léopard d'or.
 Lee H H B (1989). *Korean grammar*. Oxford: Oxford University Press.
 Li J-M (1985). *Grammaire du coréen: vol.1*. Paris: P. A. F.
 Li J-M (1991). *Grammaire du coréen: vol. 2*. Paris: P. A. F.
 Martin S E (1954). *Korean morphophonemics*. Baltimore, MD: Linguistic Society of America.
 Martin S E (1992). *A reference grammar of Korean: A complete guide to the grammar and history of the Korean language*. Tokyo: Tuttle.
 Ramstedt G J ([1939] 1979). *A Korean grammar*. Atlantic Highlands, NJ: Humanities Press.
 Sohn H-M (1994). *Korean*. London: Routledge.
 Song S C (1988). *Explorations in Korean syntax and semantics*. Berkeley, CA: University of California at Berkeley, Centre for Korean Studies.
 Yeon J (2003). *Korean grammatical constructions: their form and meaning*. London: Saffron Books.

Krio

P M K Thomas, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

The Sociohistorical Development of Krio

Krio, one of the languages spoken in Sierra Leone in West Africa, is a creole language belonging to the Atlantic group of English creoles which are restructured languages with English as the superstrate language and varying degrees of structural influences of the Niger–Congo languages in Africa. It can be further subcategorized as West African on the basis of its areal distribution and immediate linguistic affiliations.

The Domestic and Jamaican hypotheses are two competing views that have been postulated for the origins of Krio. The Domestic Hypothesis, a corollary of monogenesis, is advocated by Hancock (1986) and other creolists. It is argued that Krio is an offshoot of an English variety of creole, which coexisted with and was influenced by a Portuguese-derived pidgin in West Africa. According to Whinnom (1965), this pidgin was related to Sabir, a Romance pidgin and Mediterranean lingua franca spoken between European and non-European sailors and traders from the Middle Ages to the early 20th century. A lesser known version of the Domestic Hypothesis is based on the view stated in E. D. Jones (1956), Berry (1959), and Peterson (1969) that contact between the local inhabitants and various groups of settlers after the founding of the 'Province of Freedom' in the peninsula of Sierra Leone in the 19th century gave rise to Krio.

Sociohistorical developments in the late 18th century and the first half of the 19th century and the attendant linguistic situation around the Sierra Leone peninsula contributed significantly to the establishment and spread of present day Krio. Henry Smeathman, a botanist, who had lived for 3 years on the Banana Island near the Sierra Leone River from 1771 to 1774, proposed the area as suitable for the establishment of an agricultural settlement populated by a free community of equal blacks and whites. Smeathman died before realizing his dream, but his idea was revived by Granville Sharp, one of the philanthropists behind the campaign to repatriate and rehabilitate emancipated African ex-slaves. Following the abolition of slavery in Britain in 1787, emancipated ex-slaves in London became destitute and created a social problem for the British government. Spitzer (1974: 9) describes their destitution and the acute need for repatriation and rehabilitation of these ex-slaves who became known as the Black Poor. A group of philanthropists campaigned vigorously for the identification, purchase, and establishment

of a settlement for the Black Poor. Sharp and a group of social reformers known as the The Clapham Sect pioneered the repatriation of 411 Black Poor and some English women to the peninsula of Sierra Leone in 1787.

When they arrived in Sierra Leone, Captain Thompson, who was the leader of the expedition, purchased a piece of land from King Tom, the Temne Chief. The piece of land became known as The Province of Freedom and Thompson called the first settlement Granville Town after Granville Sharp. Historians have attributed the collapse of the first settlement to problems ranging from the fact that the settlers were ill-equipped for the weather to hostilities between the settlers and the Temnes. This occasioned a dispersal of the first settlers so that by 1791 only 48 of the Black Poor remained in The Province of Freedom.

The formation of the Sierra Leone Company and the repatriation of 1131 Africans from Nova Scotia in March 1792 saw the revival of the settlement and the return of some of the first settlers. The Nova Scotians were ex-slaves from the American colonies who had gained their freedom by fighting on the side of the British during the American War of Independence. After the war, they were offered asylum in the British settlement in Nova Scotia. When Thomas Peters, one of the ex-slaves, complained of their ill treatment in Nova Scotia, the British government transported them free to Sierra Leone where, together with the surviving Black Poor, they began a colony they called Free Town. Lieutenant John Clarkson of the Royal Navy, who led the Nova Scotians, became the first Governor of Sierra Leone. Like Granville Town, Free Town (Freetown became the capital of Sierra Leone) also had its fair share of misfortunes. Forty of the Nova Scotians died in the first few weeks of their arrival in the colony. Furthermore, the French razed the settlement to the ground in September 1794, but the surviving settlers rebuilt it.

In 1800, 550 Maroons from Jamaica arrived in the settlement. The Maroons were descendants of slaves who were originally from the Gold Coast but had been taken to Jamaica in the West Indies. They had organized several rebellious campaigns against the British and were promised an amnesty if they surrendered. The British expelled them to Halifax in Nova Scotia but during the bitter winter of 1796–1797, they petitioned to be removed to another place and were taken to Sierra Leone. Instead of being resettled in the Banana Islands, south of the peninsula, they settled in the colony after assisting to foil an uprising in the province. The Maroon population is reported to have dwindled after a decade because

some of them died from diseases and others migrated to their original home in the Gold Coast, now known as Ghana.

When the slave trade was prohibited in 1808, Freetown became a Crown Colony. From then on it served as the springboard for British legal and naval operations aimed at combating the slave trade along the West Coast of Africa. Between 1808 and 1864, slave ships were intercepted on the high seas and the redeemed Africans, referred to as Liberated Africans or Recaptives, were released initially in Freetown, and later in other British ports. The Liberated Africans never landed in the New World and, according to Spitzer (1974: 10), “were of heterogeneous ethnic origin, speaking a Babel of African languages.” The last batches of settlers were from ethnic tribes in Sierra Leone and other West African countries including Ghana and Nigeria.

The languages spoken in Freetown, including a dialect of West African pidgin/creole, coexisted with English, which was the official language used in the administration of the settlement. Even if historians and creolists disagree on when an English pidgin was first used on the coast of West Africa, or on the ethnolinguistic and demographic composition of the settlement, it is argued here that all of the languages spoken in the settlement, approximately 150 according to Koelle (1854), played roles in the creolization process that produced Krio. The African languages included Temne, the language of the people who sold The Province of Freedom to the first settlers, Mende, Sherbro, Joloff, Bambara, and Kissi. The adstratal influence of settlers from Barbados on the development of Krio probably occurred later, between 1819 and 1896 when, as Berry (1959: 299) points out, convicts from Barbados and disbanded troops from the 2nd and 4th West Indian Regiments were among the early colonists in Sierra Leone from the Caribbean area.

Thus, two periods can be identified in the development of Krio: the pre-1787 period during which a variety of West African creole was spoken in the Sierra Leone estuary and the post-1787 period (between 1787 and the 1860s) during which what is Modern Krio became established out of linguistic input from West African creole, New World creoles, and West African languages. From 1787 onward, the Sierra Leonean variety of West African creole and nascent Krio later converged as a result of pressure from and prestige of the latter. The process, which involved different creole varieties and West African languages, aptly demonstrates the roles of leveling and reconstruction in the creolization model.

Krio spread from Freetown to the interior of Sierra Leone and other West African countries due to the

strategic role of Sierra Leone as the base from which Britain spread its colonial administration. The British employed educated Krio and sent them as administrators to other West African colonies in the 19th century. Varieties of Krio are spoken today in the Gambia, Cameroon, Guinea, Senegal, Ghana, Nigeria, and Fernando Po, which is now known as the Island of Bioko.

‘Creo’ and ‘Creole’ have been used as variant names for Krio but Fyle and Jones (1980) and Wyse (1980) argue that ‘Akiriyo,’ a Yoruba term which refers to the Krio habit of paying visits after religious services is the most plausible derivation of Krio. According to Hancock (1969: 19) in the past, ‘Creole’ was the generic term for the settlers and their descendants, from 1787 to the second half of the 19th century, and their language. At an orthographic conference held in Freetown in April 1984, participants recommended the adoption of Krio as the official designation of the language and name of the people. Krio is today one of the national languages of Sierra Leone. It is estimated that native speakers of Krio constitute 3% of the population of Sierra Leone and two-thirds of the rest of the population of the country use the language as a lingua franca. The native speakers inhabit mostly Freetown and the western area of the country.

Krio has not decreolized for a variety of reasons. It has been used by writers as a medium of poetry, drama, and short stories. There is a Krio–English dictionary, portions of the Bible including the New Testament have been translated into Krio and many plays performed in Sierra Leone today are written in Krio. Most importantly, Krio has enjoyed an enhanced social status as one of the official languages of Sierra Leone. It is regarded as a full-fledged Sierra Leonean language and not a corrupt bastardized version of English. Along with the other Sierra Leonean languages, it is used in television broadcasts, as well as most business and official public engagements.

Official recognition of the language also extends to education. In the past, an additive bilingual program of language shelter was widely practiced in primary schools throughout the country. Children in primary schools used to be taught to a large extent through the medium of their first language and English was slowly introduced. The program of immersion had the educational aim of enriching the experiences of the children. The indigenous languages and cultures were maintained and further developed as they interacted with English. Sierra Leone’s current education policy requires the extension of the use of Krio and the other indigenous languages to secondary education and there is ongoing work to standardize these languages and produce materials for use at this

level. This policy will contribute to the extension of the domains of use and the continued stability and spread of Krio as well as the other national languages.

In addition to the other reasons given above, Krio has not decreolized notwithstanding pressure from English, the lexifier language, because it carries no stigma and is an identity marker.

Some Grammatical Features of Krio

Krio uses separate morphemes to express grammatical categories. In general, the plural is marked by the particle *dem*, for example:

dem pen
PLURAL pen
pens
pen dem
pen PL(URAL)
pens

Whereas English plural forms can be indicated by the suffixes *-s*, or *-es*, some Krio words appear to have suffixes attached without necessarily indicating plurality. Such words were acquired from English in their present forms and can be used with reference to both singular and plural objects, for example:

machis
 'match'
sus/shuz
 'shoe'

The possessive is marked by the particle *in*, for example:

Patrik in buk
Patrick POSSESSIVE book
 'Patrick's book'

De or *di* is the progressive marker, for example:

Idit de rait.
Edith PROGRESSIVE write.
 'Edith is writing.'

In certain contexts, verbs without preverbal particles express the default past tense but the particle *bin* is the past tense marker, for example:

Banadet bin rait.
Bernadette PAST rait.
 'Bernadette wrote.'

Go is the future marker, for example:

Lamin go rait.
Lamin FUTURE rait.
 'Lamin will write.'

Dɔn is the perfective marker and it combines with *bin* to express the past perfect, for example:

Angela dɔn rait.
Angela PRESENT PERFECT write.
 'Angela has written.'
Mamie bin dɔn rait.
Mamie PAST PERFECT write.
 'Mamie had written.'

Blan(t) 'used to' is a habitual aspect marker and the modal markers which can cooccur in a sequence as double or multiple modals are *mɔs* 'must,' *fɔ* 'should,' *go* 'intend to, must,' and *kin* 'can, could.' Consider the following examples:

I blant rait.
She HABITUAL ASPECT write.
 'She usually writes.'
A bin fɔ mɔs dɔn rait.
I PAST + MODAL + MODAL + PERFECTIVE
write.
 'I should (would) most certainly have written.'

Some particles are multifunctional and the different functions are determined by context. These particles include:

- *Na* can function as a locative 'in, at, to,' verbal particle 'is,' and adjectival 'that.'
- *ɔt*, a preposition 'out' also functions as verb for 'extinguish, put out.'
- *De* functions as a verb 'to be,' durative marker, and locative adverb 'there.'
- *Bin* functions as a past tense marker and aspect marker.
- *Fɔ* functions as a preposition, modal auxiliary, main verb, and complementizer or infinitive marker.
- *Go* functions as a verb and modal particle.
- *Blant* is not only a past habitual marker 'used to,' but also functions as a main verb 'belong to.'

Other grammatical features of Krio include:

- multiple negation, for example:

Nɔ tɔk tu am no moh.
NEGATOR talk to him NEGATOR more.
 'Do not talk to him any more (again).'

- Serial verbs in Krio have the following basic structure:

$NP_1 \text{ Aux } V_1 (NP_2) V_2 \dots$
Agnes bin kuk res gi Josef.
Agnes (NP₁) bin (PAST AUXILIARY) cook (V₁) rice
(NP₂) gi (V₂) Joseph.
 'Agnes cooked some rice which she gave to Joseph.'

- focus constructions involving *na* ‘it-is,’ for example:

Na Mari bin kɔˈl.
It-is Mary PAST call.
 ‘Mary called.’

Some Lexical Features of Krio

Many Krio words are of African origin. The main Sierra Leonean sources of the lexical items include, Mende, Temne, Sherbro, Susu, Yalunka, Limba, Kru, Vai, Fullah, and Mandingo. The other major African sources include, Yoruba, Wolof, Twi, Hausa, and Ibo. Most of the African items have multiple origins and indicate the multiple connections of the language as shown below:

| Word | Sources | Meaning |
|-----------------------------|--|---|
| <i>banda</i> | Mende, Kongo, Swahili | a basket made of palmetto straw or of marsh grass and sewn with palmetto, or a thatched house |
| <i>bara</i> (bala, balanji) | Bambara, Susu, Mandingo | a xylophone |
| <i>bene</i> | Wolof, Bambara, Mende | benne, sesame |
| <i>fufu</i> | Twi, Ewe, Wolof, Fon, Mende, and Hausa | mush, wheat flour made into a thin batter and cooked to eat, eat up, food |
| <i>nyam/nyamnyam</i> | Wolof, Fullah, Mbundu, Mandingo, Tshiluba, Efik, and Twi | |

Other sources of the lexicon of Krio include European languages: mainly English, Portuguese, Spanish, and French. The largest number of words in the lexicon of this creole derives from English. Between 70% and 80% of the Krio lexicon is derived from English and a lot of the words reflect an archaic usage in English. Examples include:

berin ‘a funeral, a burial.’ Recorded in the OED as *bering(e)*, it is either obsolete or occurs in English dialects.
titi ‘girl’ and *krabit* ‘miserly, mean’ are Scottish representations
bre/ribr ‘nag’ comes from northern English dialect
kostament ‘customer’ appears in the OED as an obsolete word *customancel/custumaunce* in use as far back as 1386, which means ‘customary practice; custom, habit, customary gathering, frequenting’
baksay ‘buttocks’ is a fossil of an earlier English compound *baksyde*, *backside* meaning ‘the hinder or back part, the back, the rear.’

Vulgar words highlight slang and vulgar usages but some of these words have lost their European connotations. Examples include:

pis ‘urine’
switpis ‘diabetes’
pisbag ‘bladder’
pishol ‘urethra’

Some of the Krio words represent semantic Africanisms, for instance:

bif ‘meat’
fut ‘leg, thigh’
met ‘co-wife’

Although some Krio words are different from their English etyma, they are related to their etyma via a semantic change through inference, for instance:

bisin ‘to care, be concerned about’ from ‘business’
drap ‘arrive unexpectedly’ from ‘drop’
bot ‘to gang up’ from ‘both’
ton ‘penis’ from ‘stone’

Words derived directly from Portuguese, Spanish, and French include:

pikin ‘child’ (derived from Portuguese *pequenno* ‘little’ or Spanish *pequeño*),
boku ‘plentiful, abundant’ (derived from French ‘*beaucoup*’),
farinha ‘flour’ (derived from Portuguese but can also be traced to French *farine*)
sabi ‘skill, knowledge’ (derived from Portuguese *sabir* ‘know’ or Spanish *saber*. It also occurs in English as *savvy*)
plaba ‘quarrel’ (derived from Portuguese *palavra* ‘words, talk,’ Spanish *palabra*, Italian *parola*, French *parole*, and Latin *parabola* ‘parable.’)
dash ‘present’ (derived from Portuguese *das-me* ‘give me’)

Compounds

Some Krio compounds are created by juxtaposition of words of different grammatical categories. Some of the compounds are derived from English phrasal verbs for example:

fɔdɔm ‘fall, fall down’
mekes ‘hurry’
tayup/tringup ‘tie up’

Two or three morpheme parallels in question words include:

wetin ‘what’
udat ‘who’
usay/wisay ‘where’
ustem/wataym ‘when’
wetin-du/mek-so ‘why’
ɔmɔs/hɔmɔɔch ‘how many’

Other compounds with different bases are instances of the use of metaphoric language through idiomatic calquing, for example:

| Adjectives | Body parts (nouns) | Compound |
|-------------------|-----------------------|-------------------------------|
| <i>big</i> 'big' | <i>mɔt</i> 'mouth' | <i>bigmɔt</i> 'boastfulness' |
| <i>gud</i> 'good' | <i>bele</i> 'belly' | <i>gudbele</i> 'kind hearted' |
| <i>big</i> 'big' | <i>yay</i> 'eye' | <i>bigyay</i> 'greedy' |

Other socially and culturally determined compounds include:

santem (sun time) 'midday'
domɔt (door mouth) 'door'
dede-hos (dead house) 'mortuary'
simun 'menstruation'

Some compounds are gender, occupation, actor, and nationality constructions, for example:

umanfɔl 'hen'
bɔy pikin 'male child'
gyal pikin 'female child'
inglishman 'Englishman'
amerikinman 'American'
ganaman 'Ghanaian'

There is also evidence of semantic calques and extensions as the forms borrowed from English gain new meaning. For example:

ɔpstyas 'the brain' (from 'upstairs')
bizi 'menstrual period' (from 'busy')
yad 'toilet' (from 'yard')
big 'older, wealthy, important'

Epenthetic vowels are inserted between consonants where clusters occur in English words, for example: *tikitul* 'kettle'

Reduplication

The following types of complete reduplication have been attested in Krio:

- intensive reduplication of adjectives, adverbs and verbs

| Simplex forms | Reduplicated morphemes |
|-----------------------|--------------------------------------|
| <i>tru</i> 'true' | <i>trutru</i> 'very true' |
| <i>de</i> 'there' | <i>dede</i> 'exactly there; correct' |
| <i>ay</i> 'high' | <i>ayay</i> 'very high' |
| <i>kwik</i> 'quickly' | <i>kwikkwik</i> 'very quickly' |

The following reduplicated verbs also indicate an increase in degree and/or intensity:

| Simplex forms | Reduplicated morphemes |
|-----------------------------|-----------------------------------|
| <i>baŋ</i> 'bang' | <i>baŋbaŋ</i> 'very loud noise' |
| <i>fred</i> 'be frightened' | <i>fredfred</i> 'very frightened' |

- iterative/repetitive/frequentative reduplication of verbs

| Simplex forms | Reduplicated morphemes |
|-----------------------|---|
| <i>aks</i> 'ask' | <i>aksaks</i> 'repeated asking around' |
| <i>chenj</i> 'change' | <i>chenjchenj</i> 'habitually/always/constantly changing' |

- distributive reduplication of numerals

| Simplex forms | Reduplicated morphemes |
|------------------|---|
| <i>wan</i> 'one' | <i>wanwan</i> 'one by one' or 'one to each' |
| <i>tu</i> 'two' | <i>tutu</i> 'two by two' or 'two to each' |

- pluralizing reduplication of nouns.
- plurality bordering on uncountability, for example:

| Simplex forms | Reduplicated morphemes |
|---------------------|------------------------------|
| <i>af</i> 'half' | <i>afaf</i> 'bits', 'halves' |
| <i>chuk</i> 'thorn' | <i>chukchuk</i> 'thorns' |

- increased mass, for example:

| Simplex form | Reduplicated morpheme |
|---------------------|-------------------------|
| <i>chaf</i> 'chaff' | <i>chafchaf</i> 'chaff' |

Some Phonological Features of Krio

Consonants

The voiceless dental fricative /θ/ and the voiced dental fricative /ð/ are often occluded. The voiceless dental fricative /θ/ is reduced to /t/ or in some dialects /f/, for example:

tin 'thin'
tiŋk 'think'

The voiced dental fricative /ð/ is replaced by the alveolar /d/, for example:

da 'that'
dis 'this'
den 'then'

The voiceless glottal fricative /h/ is often omitted in initial positions, for example:

ol 'hold'
os 'house.'

Initial unstressed sounds are omitted in many words, for example:

gri 'agree'
membra 'remember'
chenj 'exchange'
bot/bvut 'about'

Consonant clusters involving a fricative and stop in final positions are often reduced, for example:

was was/was 'wasp'
gens 'against'
han 'hand'
fɔs 'first'
bres 'breast'

The stop after /l/ is often dropped and the rule appears to be the reduction of the clusters /lt/ and /ld/ to /l/ in final positions for instance:

/sɔl/ 'salt'
 ol 'old'
 wɔl 'world'

The /l/ before a labial consonant or a dental stop at the end of a word is deleted for example:

ɛp/hɛp 'help'
 sɛf/sɛf 'self'

The final voiced stop /d/ in words ending in /nd/ is omitted, for example:

san 'sand'
 tan 'stand'
 blɛn 'blind'

Like some West African languages such as Mende and Temne, Krio has the following voiceless and voiced labiovelar coarticulated stops:

/kp/
 /gb/

Krio has the following prenasalized stops:

/mb/, /nt/, /ŋg/, and /ŋk/

Consonants are palatalized and the different types are:

- nasalized palatal glide

/ny/

- palatalized alveolars

/dy/, /sy/, /ty/, and /zy/

- palatalized velars

/gy/ and /ky/

The sound /v/ is realized as /b/ in certain words, for example:

ib 'heave'
 oba 'over'
 dreb 'drive'
 koba 'cover'

The /v/ in final positions is often rendered as /f/, for example, 'move' is pronounced /muf/.

Adjacent consonants are often transposed. This is known as metathesis and examples include:

/ask/ 'ask' is pronounced /aks/
 /risk/ 'risk' is pronounced /riks/
 /mɔsk/ 'mosque' is pronounced /mɔks/

Although this is irregular, there is evidence of rhoticism in Krio in a word like /brɛit/ 'bleach' in which /l/ is realized as /r/.

Vowels and Diphthongs

Krio has seven pure vowels and the vowels lack corresponding pairs of long and short vowels. It also has three diphthongs, /ai/, /au/, and /ɔi/. Variants of these diphthongs are /ay/, /aw/, and /oy/. Vowels are introduced to replace English diphthongs for example:

/e/ replaces /ei/
 /o/ replaces /əʊ/

Vowels are added at the end of a word. These are known as paragogic vowels. Examples include:

gladi 'glad'
 dede 'dead'
 arata 'rat'
 lili 'little'

Suprasegmentals

Krio is a tone language whose intonation is influenced by African tonal languages. There are two tones, a low tone /ˊ/ which is low in all positions and a high tone /ˋ/. The falling pitch /ˋ/ is used as a realization of the high tone. Tone is syllable-timed and each Krio word or segment has a relevant pitch pattern. Every syllable has a separate tone or relative pitch that is unrelated to stress. The pitch of syllables causes differences in meaning.

Bibliography

- Alleyne M C (1980). *Comparative Afro-American: an historical-comparative study of English-based Afro-American dialects of the New World*. Ann Arbor: Karoma.
- Berry J (1959). 'The origins of Krio vocabulary.' *Sierra Leone Studies* 12, 298–307.
- Bradshaw A T Von (1966). 'A list of Yoruba words in Krio.' *Sierra Leone Language Review* 5, 61–71.
- Coomber M E A (1969). *A descriptive study of Krio phonology*. M.S. Thesis. Georgetown University.
- Fyle C N & Jones E D (1980). *A Krio-English dictionary*. Oxford: Oxford University Press/Freetown:University Sierra Leone Press.
- Greenberg J H (1966). *The languages of Africa*. The Hague: Mouton/Bloomington: Indiana University Press.
- Hancock I F (1969). 'A provisional comparison of the English-based Atlantic Creoles.' *African Language Review* VIII, 2–72.
- Hancock I F (1986). 'The domestic hypothesis, diffusion and componentiality: an account of Atlantic anglophone creole origins.' In Muysken P & Smith N (eds.) *Substrata versus universals in creole genesis*. Papers from the Amsterdam Creole Workshop, April 1985. Amsterdam/Philadelphia: John Benjamins. 71–102.
- Holm J A (2000). *An introduction to pidgins and creoles*. Cambridge, New York: Cambridge University Press.

- Huber M & Parkvall M (eds.) (1999). *Spreading the word: the issue of diffusion among the Atlantic creoles*. London: University of Westminster Press.
- Hymes D (ed.) (1971). *The pidginization and creolization of languages*. Cambridge: Cambridge University Press.
- Jones E D (1956). 'Some aspects of the Sierra Leone patois or Krio.' *Sierra Leone Studies New Series* 6, 97–109.
- Jones E D (1968). 'Some tense, mode and aspect markers in Krio.' *African Language Review* 7, 86–89.
- Jones E D, Sandred K I et al. (eds.) (1992). *Reading and writing Krio: proceedings of a workshop held at the Institute of Public Administration and Management, University of Sierra Leone, Freetown, 29–31 January, 1990*. Uppsala: Almqvist & Wiksell International.
- Jones F C V (1983). *English-Derived Words in Sierra Leone Krio*. Unpublished Ph.D. thesis, University of Leeds.
- Koelle S (1854). *Polyglotta Africana*. Fourah Bay College, University of Sierra Leone. Reprinted 1963.
- Kup A P (1961). *A history of Sierra Leone 1400–1787*. Cambridge: Cambridge University Press.
- Ladefoged P (1964). *A phonetic study of West African languages*. Cambridge: University Press.
- Le Page R B (1960). 'Creole Language Studies I.' London: Macmillan.
- Le Page R B (1961). 'Creole Language Studies II.' London: Macmillan.
- Mufwene S S (2001). *The ecology of language evolution*. Cambridge, New York: Cambridge University Press.
- Peterson J E (1969). *Province of Freetown: a history of Sierra Leone, 1787–1870*. Evanston: Northwestern University Press.
- Romaine S (1988). *Pidgin and creole languages*. London: Longman.
- Spencer J (ed.) (1971). *The English language in West Africa*. London: Longman.
- Spitzer L (1974). *The creoles of Sierra Leone: responses to colonialism 1870–1945*. Madison: University of Wisconsin Press.
- Thomas P (2004). *Some syntactic and lexical affinities between Krio and Gullah*. Ph.D. diss., University of Oxford.
- Tinelli H (1981). *Creole phonology*. The Hague: Mouton.
- Welmers W E (1973). *African language structures*. Berkeley: University of California Press.
- Whinnom K (1965). 'The origin of European-based creoles and pidgins.' *Orbis* 14, 509–527.
- Wyse A J G (1980). *Search light on the Krio of Sierra Leone: an ethnographical study of a West African people*. Freetown: Institute of African Studies University of Sierra Leone.
- Wyse A J G (1989). *The Krio of Sierra Leone: an interpretive history*. London: Hurst in association with the International African Institute.

Kru Languages

J Bendor-Samuel, Summer Institute of Linguistics, High Wycombe, UK

© 2006 Elsevier Ltd. All rights reserved.

Kru languages are spoken mainly in the forest areas of southwestern Ivory Coast and southern Liberia. Apart from three languages, they form a contiguous block with Kwa languages to their east, Mande languages to their north and west, and Atlantic languages to their west.

Speakers

Reliable population figures are hard to obtain, but it would seem that there are approximately 2 million people who speak one or other of the Kru languages. The three largest language groups are the Guere complex (some 400 000 speakers), the Bete complex (approximately 350 000 people), and the Bassa (over 250 000 people).

Kru Studies

Though Sigismund W. Koelle, in 1854, included five Kru languages in his *Polyglotta Africana*, there was little study of these languages until this century. In

1905, Georges Thomann published a grammar of Noyo, but the next substantial work on Kru languages did not appear until 1966, when Gordon Innes published his *Introduction to Grebe* and a Grebe-English dictionary. Subsequently, however, research has begun into many more of the Kru languages, and the work of Marchese (1983, 1986, 1989) has been extensive.

Classification

Greenberg (1963) tentatively included Kru in his Kwa branch of Niger-Congo, but this classification has been rejected by most scholars. Some lexicostatistical studies and the presence of noun class suffixes have suggested that Kru is closer to Gur than to Kwa, though, unlike Gur and Kwa, Kru is nonserializing. More recently, the hypothesis of grouping Kru, Gur, and Adamawa-Ubangi as North Volta-Congo has found increasing acceptance (Williamson and Blench, 2000). However, since systematic lexical study is lacking and in the absence of any conclusive evidence, it seems best to regard Kru as an independent group for the time being.

The Kru languages can be divided into two main groups: eastern and western. Eastern Kru

(approximately 500 000 speakers) is spoken exclusively in the Ivory Coast and contains two major subgroups, the Bete complex and the Dida complex. Bete further divides into eastern Bete, spoken in the region of Gagnoa, and western Bete, spoken in a wider area, including Soubre, Guibéroua, Issia, and Daloa. The Dida complex divides into eastern Dida, with six dialects, and western Dida, with two dialect clusters. Kwadia, Bakwe, and Wane are three other eastern Kru languages that do not belong to either the Bete or the Dida complex.

Western Kru (with more than 1 million speakers) is spoken over a considerably larger area compared to eastern Kru and extends from the Sasandra River in the Ivory Coast through southern Liberia. The major division in western Kru is between the Guere complex, Grebo complex, and Bassa. The Guere complex is located in Ivory Coast and comprises some 35 languages/dialects; these may be grouped into four groups: Nyabwa, Wobe, Guere-Krahn, and Konobo. The Grebo complex straddles the Ivory Coast and Liberia border and includes some 25 languages/dialects. In the Ivory Coast, there are two main subgroups: Tepo-Plapo and Pie. In Liberia, there are seven subgroups: Wedebo, Glebo, Jabo, Gedebo, Niaibo, Fopo, and Chedepo. In Liberia, Bassa, with some 15 dialects and some 250 000 speakers, and Klao, with two main dialects, belong to western Kru but lie outside Grebo and Guere.

Three other Kwa languages, Kuwaa, Aizi, and Seme, are not grouped with either eastern or western Kru. All three are separate geographically from the rest of the Kru languages. Seme is of particular interest since it is located in Burkina Faso, over 300 miles from the nearest Kru language and surrounded by Gur languages. **Figure 1**, based on work by Marchese, gives a useful summary.

Structural Features

Phonetics and Phonology

Open syllables predominate and words are usually monosyllabic or disyllabic. Vowel sequences occur

as do CLV syllables (consonant-L-vowel, where L represents a syllabic l, n, or r). All Kru languages have stops at five points of articulation: bilabial, alveolar, palatal, velar, and labio-velar. A typical Kru vowel system has four front and four back vowels and a central vowel and is marked by vowel harmony with advanced and retracted sets of vowels. Within any one morpheme, only vowels of one set occur. Some Kru languages, including Grebo and Krahn, have another type of vowel harmony, with vowels divided into three sets, A, B, and C. Vowels of any one group may cooccur in a morpheme, and vowels of adjacent sets may cooccur. So vowels from set A may occur with vowels from set B, or vowels from set B may occur with vowels from set C, but vowels from set A never cooccur with vowels from set C. Set A consists of i, u, e, and o; set B consists of ε, a, and ɔ; and set C consists of ɪ and ʊ.

All Kru languages are marked by three or four levels of register tone, which may carry either lexical and/or grammatical functions. Tone, for instance, may distinguish the imperfective from the perfective and the singular from the plural.

Grammar and Syntax

Kru languages have a subject-verb-object (SVO) basic word order, but when an auxiliary is present this changes to S AUX (IO) (DO) V. Kru languages are suffixing. Plurality is often indicated by an *-i* suffix. A number of Kru languages have remnants of noun class suffixes and a few have noun class concord, with agreement between the head and its modifiers. Possessives precede nouns, and body parts function as postpositions indicating direction. Genitives precede nouns, but most modifiers follow their heads.

Kru languages have a basic aspectual system with imperfective and perfective forms marked by suffixes. Progressive and perfect forms are often marked by an auxiliary and so too are conditionals and potential futures. Negation is also frequently signaled by an auxiliary or a particle, though in some languages tone and word order are involved.

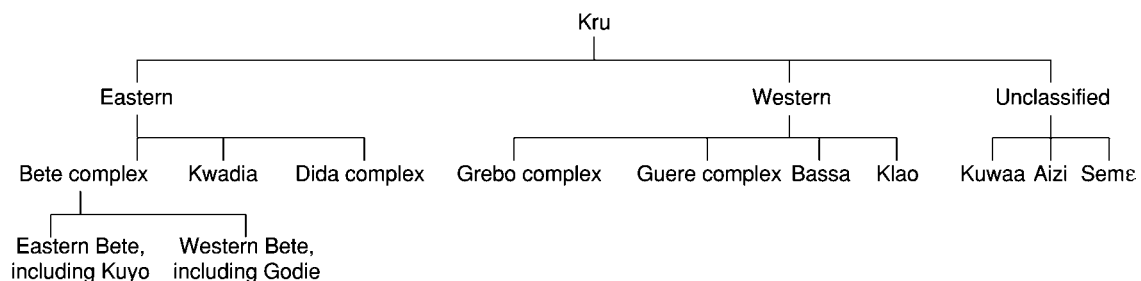


Figure 1 The Kru languages.

Bibliography

- Greenberg J H (1963). *Languages of Africa*. The Hague: Mouton & Co.
- Innes G (1966). *Introduction to Grebo*. London: School of Oriental and African Studies, University of London.
- Marchese L (1983). *Atlas linguistique kru*. Abidjan: Agence de Coopération Culturelle et Technique, Institut de Linguistique Appliquée, Université d'Abidjan.
- Marchese L (1986). *Tense/aspect and the development of auxiliaries in Kru languages*. Dallas: Summer

- Institute of Linguistics and the University of Texas at Arlington.
- Marchese L (1989). 'Kru.' In Bendor-Samuel J (ed.) *The Niger-Congo languages*. Lanham: University Press of America, Inc. 178–215.
- Thomann G (1905). *Essai de manuel de la langue néoulé*. Paris.
- Williamson K & Blench R (2000). 'Niger-Congo.' In Heine B & Nurse D (eds.) *African languages*. Cambridge: Cambridge University Press. 24–25.

Kurdish

P O Skjærvø, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

The Kurdish language belongs to the Iranian language family. It is spoken mainly in eastern Turkey, Syria, northern Iraq, western Iran, and Central Asia. Today there are large communities of Kurds living in the diaspora, for instance, Germany and Scandinavia. Kurdish is spoken in three main variants: Northern Kurdish, comprising Kurmanji in the west and dialects spoken from Armenia to Kazakhstan; Central Kurdish, spoken in northeastern Iraq (called Sorani) and adjacent areas in Iran (called Kordi or Mokri), as well as in Iranian Kurdistan (called Senne'i); and Southern Kurdish, spoken in Kermanshah province in western Iran (including Lakki and Lori of Posht-e Kuh). Northern and Central Kurdish developed rich literatures from the early 20th century on.

The earliest grammar and vocabulary of Kurdish were prepared by the Catholic missionary Maurizio Garzoni and published in 1787; these were used in subsequent scholarly descriptions of Kurdish. The earliest modern 20th-century studies were those of Oskar Mann and Karl Hadank. The first important post-World War II study, applying modern linguistic methods, was that of D. Neil MacKenzie (1961–1962).

Literary Kurdish in Iraq, Iran, and the former Soviet Union is written in the Arabo-Persian script, but employs a circumflex accent ^ placed above or below letters to mark non-Persian sounds: above *w* and *y* it denotes *majhul* vowels (*ō*, *ē*, contrasting with *ū* and *ī*), above *l* it denotes *l̥*, and below *r* it denotes the rolled *r* as opposed to the single-flap *r*. Kurdish in the USSR was also written in the Cyrillic script, with the addition of several signs and diacritics. Kurmanji Kurdish is today written in the Latin alphabet with Turkish

orthography, thus <c>=j, <ç>=č, <ş>=š, <j>=ž. In addition, a circumflex accent denotes long vowels and an umlaut on <x> denotes the voiced ɣ. In this script, <e> and <a> represent the vowels commonly transcribed as *a* [a > ɛ] and *ā* [a: > a:].

Kurdish belongs to the Central Iranian language group and, as such, has *s*, *z* from *č, *ʃ* (e.g., *āsin* 'iron', but Persian [Farsi] *āhan*; NK *az* 'I', but Old Persian *adam*). Like the Central dialects and, for instance, Parthian, it has -ž- from *-č-, -j- (*rož* ← **rau-čah* 'day', Persian *ruz*; *dirēž* ← **drājah* 'long', Persian *derāz*). It has also retained the Middle Iranian *majhul* vowels *ō*, *ē*, which in Persian have merged with *ū* and *ī*. A rare feature is the development of intervocalic *m* > *v* (including *m* < *hm* < *šm*), which in Northern and Central Kurdish remains distinct from *w*, but in Southern Kurdish merges with it (demonstrative pronouns: NK *av* 'this' ~ *aw* 'that' ← **ima*- ~ **awa*-; CK both *aw* except an area with *am* ~ *aw*; SK *ī* (cf. Persian *īn*) ~ *ow*; NK *čāv*, CK *čāv* 'eye'; CK *awa* ← **ašmā* 'you'). Kurdish shares with Persian the development of **w*- → *b*- (*bā* 'wind' ← **wātah*, Persian *bād*). Northern Kurdish retains final -*t*, which elsewhere becomes *d* or is lost (*dīt* 'saw', Persian *did*).

The Kurdish dialects have very complex phonologies, morphologies, morphophonologies, and syntax, of which no idea can be given in a small space. Proto-Kurdish had two genders, masculine and feminine, two numbers, and two cases, direct and oblique, as well as a vocative. These are preserved in Northern Kurdish, but gradually merge into a no-gender, no-case system as one moves southward (cf. the 1st sing. personal pronoun 'I, me': NK DIR *az* ~ OBL *min*, CK (*a*)*min*, SK *amin*). All three groups have an indefinite suffix going back to *-*ek* 'one', while Central and Southern Kurdish have a definite suffix going back to *-*aka* (-*ak*). The 3rd singular enclitic pronoun is -*ī*, pl. -*yān* (cf. Persian -*eš*, -*ešān*). The 1st and 2nd plural

enclitic pronouns are archaic: 1PL *-n*, 2PL *-û* (beside *-mān* and *-tān*), which go back to Old Iranian **nah* and **wah*. The ezafe has two genders and two numbers, but is simplified according to the general tendency.

The verbal systems are of the typical modern Iranian type, with a three-stem system: present-imperfect, past, and perfect, as well as a split ergative. The present and imperfect take the prefixes NK *t-/di-*, CK *a-/da-*, SK *a-/mi-* (cf. Persian *mi-*) to express progressive tense. Present forms without a prefix or with the prefix *bi-* are subjunctive. The past tenses are ergative, but only Northern Kurdish has the pure passive type construction (1).

- (1) ta az nās kir-im
 you.SING.OBL I.DIR familiar do.PAST-1ST.SING
 ‘you knew me’

In Central and Southern Kurdish, where there is no case distinction, the agent is expressed by enclitic pronouns. If the verb has a preverb, the enclitic pronouns come after it (2), otherwise they come between the verb stem and the ending (3).

- (2) a-t xwārd
 PROG-you.AGT eat.PAST[-3SG]
 ‘you were eating’

- (3) nārd-it-în
 send.PAST-you.AGENT-COR.1PL
 ‘you sent us’

Note constructions of the type in (4), where the apposition *pēwa* ‘about’ governs the person implicit in the copula *-ît* ‘you are’; (5), where ‘to’ governs the pronoun implicit in the copula *-n-* ‘you are’; and (6), where ‘to’ governs the pronoun implicit in the copula *-m-* ‘I am’.

- (4) xaw-im pēwa dīw-ît
 dream-1SG.ENCL about see.PAST-COP.2SG
 ‘I dreamed about you’

- (5) dā-m-î-n-ê
 give.PAST-COP.1SG-he.ENCL.OBL-COP.2PL-to
 ‘he gave me to you’

- (6) kitēb-ek-ān-it dā-m-ê
 book-DEF-PLUR-you.AGT give.PAST-COP.1SG-to
 ‘you gave me the books’

In Northern Kurdish, the passive is formed with the auxiliary *hāt* ‘come’ + INF (7).

- (7) hāt-īye girt-in
 come.PERF[-3SG] seize.PAST-INF
 ‘has been seized’

Central and Southern Kurdish have passive formations in present *rê-* or *yê-*, past *rā-* or *yā-* (*kuž-yā* ‘be killed’; *de-nūs-rê* [PROG-write.PRES-PASS.PRES. [-3SG]] ‘it is being written’).

Derivational nominal suffixes are common, as are a variety of types of compound. The meaning of verbs can be modified by preverbs, of which there are many, or the postverb *-awa/-āwa* (*-āw*), unique to Kurdish among Iranian dialects (*hāt-in* [come-INF] ‘to come’, but *hāt-in-āw* ‘to come back, return’). Verbal idioms consisting of adjectives or nouns plus verbs are common, as in all Iranian languages.

Bibliography

- Blau J (1989). ‘Le kurde.’ In Schmitt R (ed.) *Compendium linguarum iranicarum*. Wiesbaden: Reichert. 327–335.
 Blau J (2000). *Méthode de kurde: sorani* (rev. edn.). Paris: Harmattan.
 Blau J & Barak V (1999). *Manuel de kurde: kurmanji*. Paris: Harmattan.
 MacKenzie D N (1961). ‘The origins of Kurdish.’ *Transactions of the Philological Society* 59, 68–86.
 MacKenzie D N (1961–1962). *Kurdish dialect studies* (2 vols.). London/New York: Oxford University Press.
 MacKenzie D N (1986). ‘Kurds, Kurdistan v. language.’ In Bearman P J, Bianquis T, Bosworth C E, van Donzel E & Heinrichs W P (eds.) *Encyclopædia of Islam*, new edn., vol. 5. Leiden: Brill. 479–480.

Kurukh

P S Subrahmanyam, Annamalai University, Bangalore, India

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Kurukh is one of the tribal languages of the Dravidian family with a population of about 1.5 million; thus it is next only to Gondi within the family in terms of

number of speakers. The name of the language is spelt also as Kurukh; another name, Oraon, was also in use in earlier times. The main concentration of Kurukh speakers is found in Chota Nagpur and Bhagalpur districts of Bihar, India; some of them live also in Madhya Pradesh (Raygarh and Sarguja districts) and Orissa (Sundargarh and Sambalpur districts). Some have migrated in recent times to the tea districts of Assam and Nepal and are known there as

Dhāngar/Dhangar ‘men who receive dhān “rice” as wages’. It belongs to the North Dravidian subgroup along with Malto, its closest ally, and Brahui.

Phonology

Kuṛux contains the ten-vowel system that is normally found in the Dravidian languages (see Table 1); there are also a few words with nasalized vowels, e.g., *khē:sō* ‘blood’. Its consonant system is presented in Table 2. The velar voiceless fricative (x) and the glottal stop (ʔ) are the peculiar features of Kuṛux phonology, e.g., *xay* ‘wife’, *ciʔinā* ‘to give’.

Syntax

Word Classes

The following word classes may be recognized for Kuṛux: nouns (including pronouns and numerals), verbs, adjectives, adverbs (including expressives), particles, and interjections.

The class of adjectives is a small one. An adjective occurs before the noun it qualifies. Most of the nouns can function as adjectives, for example:

mechā partā
high, height mountain
‘high mountain’

Table 1 Vowels of Kuṛux

| | Front | | Central | | Back | |
|------|-------|------|---------|------|-------|------|
| | Short | Long | Short | Long | Short | Long |
| High | i | ī | | | u | ū |
| Mid | e | ē | | | o | ō |
| Low | | | a | ā | | |

Table 2 Consonants of Kuṛux

| | L | D | R | P | Vel | G |
|-----------|----|----|----|----|-----|---|
| Stop | | | | | | |
| VL | p | t | ṭ | c | k | ʔ |
| VLA | ph | th | ṭh | ch | kh | |
| VD | b | d | ḍ | j | g | |
| VA | bh | dh | ḍh | jh | gh | |
| Nasal | m | n | | | ŋ | |
| Fricative | | s | | | x | h |
| Lateral | | l | | | | |
| Trill | | r | | | | |
| Flap | | | ɾ | | | |
| Semivowel | w | | | y | | |

(Abbreviations: D = dental, G = glottal, L = labial, P = palatal, R = retroflex, VD = voiced (unaspirated), VA = voiced aspirated, Vel = velar, VL = voiceless (unaspirated), VLA = voiceless aspirated.)

The adjectives derived from the three deictic bases (see ‘Pronouns’) show agreement for number (but not for gender) with the noun that follows. Thus, *ā* ‘that’, *ī* ‘this’ and *ū* ‘that at a greater distance’ are used with a singular noun, e.g., *ā/ī/ū ālas* ‘that/this/that (extra-dist.) man’. The corresponding forms used with a plural noun are *abṛā*, *ibṛā*, and *hubṛā*, e.g., *abṛā/ibṛā/hubṛā āl-ar* ‘those/these/those (extra-dist.) men’. (For verbal adjectives, see ‘Nonfinite Verbs’).

An adverb occurs before the verb, for example:

ād xanem xanem bar-ckī raʔī
she again again come-past was
‘She came frequently.’

Adverbs may be divided into those of (a) time (e.g., *cērō* ‘yesterday’, *innā* ‘today’, *nēlā* ‘tomorrow’), (b) place (e.g., *mund* ‘before, in front’, *iyā* ‘here’, *ayā* ‘there’, *eksan* ‘where’), and (c) manner (e.g., *baggī* ‘much’, *oṇṭā oṇṭā* ‘separately, one by one’, *dav/daḍim* ‘well’).

The particle of emphasis is a good example for particles. It is +*m/im/am/lem*, e.g., *nēlā+m* ‘tomorrow itself’ (: *nēlā* ‘tomorrow’), *ēn-im* ‘even I’, *ār-im* ‘even they’.

Examples for interjections include: *haʔī* ‘yes’, *mal/malʔā/mallā* ‘no’, *anti(jē)* ‘of course’.

Word Order

The favored word order in Kuṛux is S(ubject) O(bject) V(erb):

ās eṅg-āgē mandar ci-c-as
he I-DAT medicine give-PAST-3M.SG
‘He gave me medicine.’

Gender and Number

Kuṛux shows a two-way distinction in gender but the classification differs in the singular and the plural; the feminine goes with the nonhuman in the singular but with the masculine in the plural, as illustrated by the following pronouns:

ās
‘he’
ād
‘she, it’
ār
‘they (human)’
abrā
‘they (nonhuman)’.

Agreement

The finite verb shows agreement with the subject pronoun (or a corresponding noun in the case of the third person) by a change in the personal

Table 3 The finite forms of *ij?* ‘to stand’

| | Pronoun | Past tense | Present tense | Future tense |
|--------------------------|---------|------------|---------------|--------------|
| 1sg. | ēn | ij-k-an | ij-d-an | ij?o-n |
| 1pl. (excl.) | ēm | ij-k-am | ij-d-am | ij?o-m |
| 1pl. (incl.) | nām | ij-k-at | ij-d-at | ij?o-t |
| 2sg. | nīn | ij-k-ay | ij-d-ay | ij?o-y |
| 2sg. (f.) | | ij-k-i | ij-d-i | ij?o-y |
| 2pl. | nīm | ij-k-ar | ij-d-ar | ij?o-r |
| 3m.sg | ās | ijj-Ø-as | ij-d-as | ij?o-s |
| 3h.pl. | ār | ijj-Ø-ar | ij-n-ar | ij?o-r |
| 3non-m.sg./ non-h.pl. | abrā | ijj-Ø-a | ij?-i | ij?o-o |

suffix (the final morpheme in the forms shown in Table 3). In equational sentences, the predicate noun shows agreement with the subject pronoun by taking the same personal suffixes as the finite verb (except 3rd non-m. sg./non-hum. pl. [see ‘Finite Verbs’]). The use of the copular verb (*tal-* ‘to be’) in such sentences is optional. Kuṛux partially preserves the old Dravidian feature of the absence of the copular verb:

ēn kuṛux-an (tal-d-an)
I Kuṛux-1sg (be-pres-1sg)
 ‘I am a Kuṛux (speaker).’
 ēm kuṛux-am (tal-d-am)
we (excl.) Kuṛux-1pl.excl (be-pres-1pl.excl)
 ‘We (excl.) are Kuṛux (speakers).’

For agreement between the demonstrative adjectives and the nouns qualified with regard to number, see ‘Word Classes.’

Noun Morphology

A nominal base is followed by the plural suffix when plurality has to be expressed; the case suffix/postposition occurs at the end. A masculine noun may take the definite suffix *-as*, e.g., *āl* ‘man’, *āl-as* ‘a particular man’.

Plural Suffixes

The plural suffix for the human nouns is *-ar*, e.g., *āl-ar* ‘men’, *mukk-ar* ‘women’ (sg. *mukkā*). Some feminine nouns take *-guṭhi-ar*, e.g., *āl-guṭhi-ar* ‘wives’; *-guṭhi (-ar)* is also added optionally to human nouns with *-ar* plural, e.g., *āl-ar(-guṭhi(-ar))* ‘men’. Kinship terms, however, take *-baggar*, e.g., *dādā-baggar* ‘elder brothers’. The word *xadd-xarrā* ‘children’ seems to contain another plural suffix *-xarrā*. Nonhuman nouns optionally add *-guṭhi*, which in origin is a

separate word meaning ‘group’, to form the plural, for example:

münd aḍḍō(-guṭhi)
three ox-PL
 ‘three oxen’
 enḍ man(-guṭhi)
two tree-PL
 ‘two trees’.

Case Suffixes and Postpositions

The nominative is unmarked. When a noun is used in the vocative, *-ō/-ay(ō)* is added to it at the end; the vocative form may be preceded by *ē/ana* (but *anay* when a woman is addressed), e.g., *ē/ana urb-ay(ō)* ‘O master!’, *anay mukk-ay* ‘O woman!’. Further, when women talk to women, *anay* is replaced by *ān/anē*, e.g., *ān xay* ‘O daughter!’, *anē xay-guṭhi-ar-ō* ‘O daughters!’.

The accusative suffix is *-an* after a consonant and *-n* after a vowel, e.g., *āl-an* ‘man (accus.)’, *mukka-n* ‘woman (accus.)’ (: *mukkā* ‘woman’) but it is *-in* after the definite suffix *-as* (in masculine nouns, e.g., *āl-as-in* ‘the man (accus.)’, the plural *-ar* (e.g., *āl-ar-in* ‘men (accus.)’) and after a demonstrative pronoun that ends in a consonant, e.g., *ād-in* ‘her, it (accus.)’.

The instrumental suffix is *-trī/-trū*, e.g., *āl-trī* ‘by the man’.

The dative uses the postposition *+gē* (variant *+ā(gē)* after a pronoun), e.g., *āl+gē* ‘to the man’, *eng-ā(gē)* ‘to me’.

The ablative suffix is *-tī* after a consonant, *-ntī* after a vowel; after the plural *-ar*, *-tī* freely varies with *-intī*, e.g., *āl-tī* ‘from the man’, *mukka-ntī* ‘from the woman’, *āl-ar-tī /āl-ar-intī* ‘from the men’.

The genitive commonly uses the postposition *+gahi*, e.g., *āl+gahi* ‘of the man’; it has the variant *+hay* after a pronoun, e.g., *eng+hay* ‘my’. However, nouns denoting a place take the suffix *-ntā*, e.g., *padda-ntā* ‘of the village’ (: *paddā* ‘village’).

The locative uses the postposition *+nū* (also *+nō* dialectally), e.g., *erpā+nū* ‘in the house’.

Pronouns

Two important features of the pronominal system of Kuṛux are the presence of inclusive and exclusive distinction in the first person pronoun (represented also in the finite verb) and the formation of the third person pronouns on three deictic bases.

The personal pronouns are:

ēn ‘I’ ēm ‘we (excl.)’
 nām ‘we (incl.)’
 nīn ‘you (sg.)’ nīm ‘you (pl.)’

The third person pronouns are formed on the three deictic bases *a-/ā-* (distant), *i-/ī-* (proximate) and *hu-/hū-* (extra-distant):

| Distant | Proximate | Extra-distant | |
|---------|-----------|---------------|-------------------|
| ā-s | ī-s | hū-s | ‘he’ |
| ā-d | ī-d | hū-d | ‘she, it’ |
| ā-r | ī-r | hū-r | ‘they (hum.)’ |
| a-bṛā | i-bṛā | hu-bṛā | ‘they (non-hum.)’ |

The reflexive pronouns are *tān* (sg.) and *tām* (pl.).

There are five interrogative pronouns: *nē* ‘who’, *ēkā* ‘who, which’, *endṛ/endrā/ēkdā* ‘what, which’. The addition of the particle *+ʔim/ʔam* to an interrogative pronoun converts it into an indefinite pronoun, e.g., *nīk+ʔim* ‘someone’.

Numerals

Kurukh retains the Dravidian numerals only up to 4 and the rest are borrowed from Hindi. The native numerals have also preserved a distinction between nonhuman and human (the variants given in parentheses for the human forms occur before case suffixes and postpositions).

| | Nonhuman | Human |
|-----|----------|----------------|
| ‘1’ | oṇḍ/oṇṭā | ort |
| ‘2’ | eṇḍ/e:ṇḍ | irb (irbar-) |
| ‘3’ | mūnd | nub (nubar-) |
| ‘4’ | nāx | naib (naibar-) |

The human numerals are generally followed by the classifier *jhan-ar* ‘persons’ (sg. *jhan*), for example:

| | | |
|-------------------|---------|-----------------|
| nub | jhan-ar | bar-c-ar |
| three (h.) | CLASSIF | come-PAST-H.PL. |
| ‘Three men came.’ | | |

The counting is done in terms of ‘score’ for which the word is *kūrī/bīsoʔe*, *oṇḍ kūrī/bīsoʔe* ‘one score’, *kūrīeṇḍ* or *eṇḍ bīsoʔe* ‘two score’, etc. The ordinal is formed by adding *-(an)-tā* to a cardinal, e.g., *eṇḍ-(an)tā* ‘second’; however, there is a special word for ‘first’: *mund-(an)tā*.

Verb Morphology

Female Speech

An important characteristic feature of the Kurukh verb morphology is the use of special forms by women when they talk among themselves; some of the tense and personal suffixes in women’s speech are different from those in men’s speech:

| | | | |
|------------------|------------------------------|---------|-------------|
| (Men’s speech) | nīn | ekātarā | kaʔa-d-ay |
| (Women’s speech) | nīn | ekātarā | kaʔa-d-ī |
| | you (sg.) | where | go-PRES-2SG |
| | ‘Where are you (sg.) going?’ | | |

Verb Bases

A verb base in Kurukh may be simple or complex. Four types may be recognized among the complex bases:

1. Intransitives or reflexives derived from transitives by the addition of *-r-*, e.g., *kol-r-* ‘to be opened’ (: *kol-* ‘to open’), *mūjh-r-* ‘to wash one’s own face’ (: *mūjh-* ‘to wash another’s face’).
2. Transitives derived from intransitives by the addition of *-aʔa-*, e.g., *maṛx-aʔa-* ‘to make dirty’ (: *maṛx-* ‘to be dirty’).
3. Causatives derived from transitives by the addition of *-tāʔa-*, e.g., *es-tāʔa-* ‘to cause to break’ (: *ess-* ‘to break’).
4. Reciprocals derived by the addition of *-nagr-* to a simple verb base, e.g., *ēr-nagr-* ‘to look at each other’ (: *ēr-* ‘to see’), *kēb-nagr-* ‘to abuse one another’ (: *kēb-* ‘to abuse’).

Finite Verbs

A finite verb has the following structure:

Verb Base + Tense Suffix + Personal Suffix

The personal suffixes are:

| | |
|---|---------------------------|
| 1st sg. | -an |
| 1st pl. (excl.) | -am |
| 1st pl. (incl.) | -at |
| 2nd sg. | -ay // (female speech) -i |
| 2nd pl. | -ar |
| 3rd m. sg. | -as |
| 3rd hum. pl. | -ar |
| 3rd non-m. sg./ -a (after past)/ -i (after present) | |
| non-hum. pl. | |

The vowel of all these suffixes is deleted after the future suffix *-o*. Three tenses, past (suffixes: *-k-*, *-ck-*), present (suffixes: *-d-*, *-n-*) and future (*-o-*) are there (see Table 3).

Unlike its counterpart in the sister languages, the imperative in Kurukh has only one form without distinction between the singular and the plural. The suffix is *-ā* when men are addressed and *-ay* when women are addressed (when women speak to women, the suffix is *-ē*): *esʔ-ā/-ay/-ē* ‘Break!’. A milder sort of the imperative has the suffix *-kē*, e.g., *bar-kē* ‘Come (if you please)!’.

Nonfinite Verbs

The infinitive, which can also function as a noun and an adjective, has the suffix *-nā* (also *-ā* in certain constructions), e.g., *es-nā/esʔ-ā* ‘to break, breaking’.

The present participle has the suffix *-nū(tī)/-num*, e.g., *es-nū(tī)/es-num* ‘breaking’.

The past participle has the suffix *-ār*, e.g., *esʔ-ār* ‘having broken’.

There are three types of adjectives derived from verb bases:

- The infinitive with adjectival function:

kūr-nā amm
burning *water*
 ‘hot water’

- Past adjective with the suffix *-ckā*:

ke-ckā āl-ar
die-PAST ADJ *man-PL*
 ‘dead people’

- Agent noun (see *Agent Noun*) functioning as nonpast adjective:

pār-ū/ pār-nā pellō
sing-NON-PAST ADJ *girl*
 ‘singing girl’

Agent Noun

The agent noun, with the suffix *-ū* can also take the masculine and the plural suffixes, e.g., (from *esʔ*- ‘to break’)

isʔ-ū ‘one who/which breaks’
 isʔ-u-s ‘a man who breaks’
 isʔ-u-r ‘persons who break’

Negative Verb

Kuṛux employs the verb base *mal-* ‘to be not’ in the present tense (with *-y-* and *-k-* as optional variants of the regular present suffixes *-d/-n-* [see Table 3]) as the negative verb to deny the identity of the subject and the predicate noun phrases, for example:

ēn bēl-an mal-d-an/mal-y-an/mal-k-an
I *king-1SG* *not.be-PRES-1SG*
 ‘I am not a king.’

Bibliography

- Bhat D N S (1966). ‘Kuṛux nouns.’ *Indian Linguistics* 27, 102–108.
- Bhat D N S (1970a). ‘Kuṛux indicatives.’ *Indo-Iranian Journal* 12, 216–223.
- Bhat D N S (1970b). ‘Velar consonants in Kuṛux.’ *Indian Linguistics* 31, 177–181.
- Grignard A (1924a). *A grammar of the Oraon language*. Calcutta: Catholic Orphan Press.
- Grignard A (1924b). *An Oraon-English dictionary*. Calcutta and Vienna.
- Hahn F (1900). *Kurukh grammar*. Calcutta: Bengal Secretariat Press. (Repr. Delhi Mittal Publications [no date].)
- Hahn F (1903). *Kurukh (Orāō)-English dictionary*. Part I. Calcutta.
- Pfeiffer M (1972). *Elements of Kuṛux historical phonology*. Leiden: E. J. Brill.

Kwa Languages

M E Kropp Dakubu, University of Ghana, Legon, Accra, Ghana

© 2006 Elsevier Ltd. All rights reserved.

The Kwa languages, a branch of the Niger-Congo family, are spoken in West Africa. As presently defined, the group extends from southeastern Ivory Coast in the west through the southern two-thirds of Ghana, southern Togo, and the Republic of Benin to the Benin-Nigeria border in the east. The term Kwa was adopted by linguists toward the end of the 19th century to group together Akan, Ga, Ewe (Éwé), and their close relatives because in many of these languages a stem *kwa* or *kua* means ‘man’ or ‘person.’

Constitution of the Group

The list of languages included in Kwa has varied considerably. In the 1950s, Westermann and Bryan included 10 languages of Ivory Coast that they

subgrouped into the Lagoon group; Akan, including the languages more recently termed Tano; Ga-Dangme; Ewe; and most of the languages of southern Nigeria. Greenberg’s 1963 classification expanded the group to include the Kru languages of Liberia in the east, the languages of the Ghana-Togo hill country that Westermann and Bryan had explicitly excluded, and the Ijaw languages of the Niger delta. This membership was maintained by Stewart in 1971, but in 1989 he eliminated the Kru languages and the Nigerian languages, keeping the Togo Mountain languages and adding a few languages hitherto unknown or unstudied. This is the membership in the most recent overview presented by Williamson and Blench (2000), which essentially has contracted back to the earliest membership plus the Lagoon and the Togo Mountain languages. The group as most recently defined, without Kru and without the Nigerian languages, is sometimes referred to as New Kwa.

Nevertheless, the Kwa languages are currently considered to be more closely related to Yoruba, Igbo, and other languages of southern Nigeria than to the rest of Niger-Congo, forming with them a branch of the Volta-Congo subgroup of Niger-Congo. However, except at the lower levels of classification such as the Tano, Potou-Tano, and Ewe-Fon (Gbe) groups, genetic relationships among these languages are quite distant. It has never been adequately demonstrated using the comparative method that Akan, Ga, Ewe, and the Togo Mountain languages are more closely related to one another than to any other languages.

The Subgroups

The (New) Kwa subgroups are distributed approximately from east to west along the coastal forest and savannah, with the majority of languages and speakers in Ghana. However, many communities have traditions of migration, and it is likely that 1500 years ago their geographical distribution was very different from today. A primary distinction has sometimes been made between the western languages with Ga-Dangme and some of the Togo Mountain languages, referred to as 'Nyo,' and Gbe and the other Togo Mountain languages, referred to as 'Left Bank' (of the Volta River). We examine the composition of the subgroups starting in the west and moving eastward.

The Lagoon languages of the Ivory Coast – Avikam and Alladian, Adjukru (Adioukru), Abidji, Abbey (Abé), Attié, and an isolated language farther west called Ega – are not particularly closely related to one another. Just east of these, the Potou-Tano group comprises at least 20 languages, including Ebré (Ebríé) and Mbatto (Mbato) (the Potou of Potou-Tano) and the Tano group consisting of Krobu, Abure, and Eotilé (Beti) (all spoken in the Ivory Coast); Nzema, Ahanta, Anyi (Anyin), and Baule (Baoule) (spoken in the eastern Ivory Coast and western Ghana); Anufo (spoken in northeastern Ghana); Akan (with by far the most speakers, mainly in Ghana); and approximately a dozen languages making up the Guang group including Gonja, Nawuri, Nchumuru (Nchumbulu), Krachi (Krache), Nkonya, Okere (Cherepon), Larteh, Efutu, and Awutu (all spoken in Ghana from north of Akan to the sea in the south, distributed along the course of the Volta river). Because the locus of greatest differentiation is in the west, it has been suggested that the Potou-Tano languages spread from west to east. Migration traditions suggest that they also spread in the eastern area (i.e., mainly in Ghana) from north to south.

Ga-Dangme, on the other hand, consists of just two languages, Ga and Dangme (or Adangme). Now situated on the Accra Plain south east of the Tano languages and west of the Volta River, the speakers have traditions of migration from the east, possibly in conjunction with the Ewe, and from the northeast, probably from what is now the northern Republic of Benin.

The Central Togo or Togo Mountain languages comprise 14 languages spoken in Ghana, Togo, and the Republic of Benin, most of them in the hilly area on the Ghana-Togo border. An earlier name for these languages was Togo Remnant Languages, adapted from the German *Togorestsprachen*. That name reflected the possibility that their fragmented distribution on hilltops difficult to access and their generally small numbers of speakers result from these linguistic communities having fled to refuge areas at some time in the past, under pressure from expanding larger groups such as (perhaps) the Ewe and the Akan. True or not, the name is disliked by speakers and no longer used.

Heine (1968) treated the languages as a genetic family and classified them into two subgroups: the NA subgroup (Basila (Anii), Lelemi-Lefana (Lelemi), Logba, Adele, Likpe (Sekpele), Santrokofi (Sele), and Akpafu-Lolobi (Siwu)) and the KA subgroup (Avatime, Nyangbo-Tafi, Bowiri (Tuwili), Ahlo, Kposo (Akposo), Kebu (Akebou), and Animere). More recent classifications accept these two groupings, but put them directly under Kwa; that is, they are now considered no more closely related to one another than to the other New Kwa groups, and 'Togo Mountain' or 'Central Togo languages' are merely geographical labels.

Kposo has the most speakers by a considerable margin, approximately 80 000, followed by Kebu with 17 500. (Both are spoken mainly in Togo; figures are based on Heine, 1968.) Lelemi-Lefana has approximately 15 000 speakers and Adele approximately 8000. Others have fewer: Logba reportedly has approximately 2000 speakers, and Animere was said to have fewer than 300 in the 1960s. There is evidence that one or two languages have died out in the area in the course of the past 2 centuries.

The area extending approximately 80 miles inland along the coast from the Volta eastward to the Nigerian border is dominated by the closely knit Gbe group of languages and dialects. The name is adopted from their common word for 'language'. They include Ewe, spoken in Ghana and Togo by more than 3 million people. The Ewe-speaking people have a tradition of migration from farther east but still within the Gbe area, from Nuatja and Tado in southern Togo. Gen (Gen-gbe) is the language of Lomé, the

capital of Togo and a major lingua franca. Another major variety is Aja (Aja-gbe). Somewhat more distantly related are Maxi (Maxi-gbe) and Fon (Fon-gbe), the language of the old kingdom of Dahomey, which is now part of the Republic of Benin. The easternmost variety is Gun (Gun-gbe), which crosses the Nigerian border along the coast and is thought to have been spoken in the Lagos area in earlier times.

Linguistic Characteristics

In all Kwa languages, the fundamental word order is subject-verb-object. It was once thought that they typically had very simple morphology, but, although the situation varies among the different groups, this is generally not really the case. Part of the reason for this belief may have been the fact that in many of the languages tone plays an important role in morphology, with some grammatical morphemes being realized only by tone. Thus, in the following Ga sentence only the high tone on the second syllable of the pronoun *amɛ* 'they' shows that the verb is perfect and not aorist.

amé-bá
'they have come'
they.PERF-come

Tone is also important in making lexical distinctions; again in Ga, we have *lá* 'blood' with high tone but *lā* 'fire' with low. The number of phonemic tones ranges from the minimal two in Ga and Akan to five in Avatime. In all the languages west of the Volta, tone levels are affected by downstepping, or the lowering of all pitch levels after a low tone throughout a sentence, but many languages east of the Volta, including Ewe, do not have this.

Cross-height vowel harmony, based on the position of the tongue root, is typical of the Tano languages but not of Ga-Dangme, Gbe, or some of the Togo Mountain languages. In all (Adjukru is an exception), syllables are typically open, ending in a vowel or sometimes a syllabic consonant.

Double articulated consonants ($[k̠p]$, less often $[g̠b]$) are typical of Ga-Dangme, the Togo Mountain languages, and Gbe but not of Akan and some other Potou-Tano languages. In Ga-Dangme and Gbe (and also Yoruba), but again not in the western languages, the voiceless bilabial stop /p/ did not exist until it was introduced through loanwords in the course of the 18th century. Consonant clusters either do not exist or the second consonant is limited to off-glides and /l/ or /r/ and the cluster is analyzable as being derived from a CVV or CVCV structure.

Verb systems vary, but paradigms generally express aspect features, with the perfective-imperfective

distinction basic. Grammatical tense is less important, often limited to a contrast between future and nonfuture, and even that shows signs of having been grammaticized relatively recently in many languages. The expression of progressive aspect is of particular interest because the western languages differ radically in its expression from many of those spoken east of the Volta. In Tano languages and also in Ga, the progressive (like most other aspect features) is shown by a prefix, as in the following sentences in Akan (1) and Ga (2).

(1) abofra no re-di nneema
child the PROG-eat things
'the child is eating something'

(2) gbeke le mii-ye nii
child the PROG-eat things
'the child is eating something'

In Ewe (3) and Dangme (4), on the other hand, it is periphrastically expressed, with an aspect verb followed by a nonfinite form of the event verb preceded by its object.

(3) Ɖevi a le nu ɖu-m
child the is thing eat-PROG
'the child is eating (something)'

(4) jukwe ɔ ɲe nɔ ye-e
child the is thing eat-PROG
'the child is eating (something)'

Some of the Togo Mountain languages use the periphrastic style of expression, but many do not.

In Kwa languages, the noun is followed by the adjective, article, and so on that modify it, whereas a possessor precedes the possessed noun. However, there are radical differences in nominal morphology. The Togo Mountain languages generally have several singular and plural classes expressed by prefixes; modifiers take prefixes that show agreement (concord) with the class of the noun they modify and pronouns also vary according to the class of the noun referred to. Akan and other Tano languages have several singular and plural prefixes for nouns but no class concord. Ga-Dangme, however, uses basically just one plural (no singular) suffix, whereas the Gbe languages use a phrase-final particle, as in the following example from Fon.

(5) xɔ̀ntɔ̀n yé̀tɔ̀n wè lé
friend your 2 PL
'both your friends'

Serial verb constructions, in which several verbs share one subject and sometimes object with no intervening conjunction, occur in all Kwa languages.

Bibliography

- Capo H C (1991). *A comparative phonology of Gbe*. Berlin: Mouton de Gruyter.
- Dakubu M E K (ed.) (1988). *The languages of Ghana*. London: Kegan Paul International.
- Dakubu M E K & Osam E K (eds.) (2004). *Studies in the languages of the Volta Basin* (vol. 2). Legon, Ghana: University of Ghana Linguistics Department.
- Heine B (1968). *Die Verbreitung und Gliederung der Togo-restsprachen*. Berlin: Dietrich Reimer Verlag.
- Herauld G (ed.) (1982). *Atlas des langues kwa de Côte d'Ivoire*. Abidjan, Côte d'Ivoire: Université d'Abidjan Institut de Linguistique Appliquée.
- Leben W R & Ahoua F (eds.) (2002). *Wortkunst und Dokumentartexte in Afrikanischen Sprachen 12: Contes et textes documentaires kwa de Côte d'Ivoire*. Cologne: Rüdiger Köppe.
- Stewart J M (1971). 'Niger-Congo, Kwa.' In Sebeok T A (ed.) *Current trends in linguistics, vol. 7: Linguistics in sub-Saharan Africa*. The Hague and Paris: Mouton. 179–212.
- Stewart J M (1989). 'Kwa.' In Bendor-Samuel J (ed.) *The Niger-Congo languages*. New York: University Press of America. 217–245.
- Westermann D & Bryan M A (1952). *Handbook of African languages part II: Languages of West Africa*. London: Oxford University Press.
- Williamson K & Blench R (2000). 'Niger-Congo.' In Heine B & Nurse D (eds.) *African languages, an introduction*. Cambridge, UK: Cambridge University Press. 11–42.

This page intentionally left blank



Lahnda

C Shackle, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

The name ‘Lahnda’ (Panjabi *lahindā* ‘western’), like the more natural feminine ‘Lahndi,’ is an invented blanket term without local currency used to distinguish the Indo-Aryan dialects spoken in western Panjab and the adjacent regions of Pakistan (by some 30 million speakers) from the Panjabi proper native to the central and eastern districts of the Panjab. The boundary between Lahnda and Panjabi (Shackle, 2003) is anyway an uncertain one. Many common features, e.g., retained historical geminates or the possessive marker *dā*, have been reinforced by overlapping literary traditions. Since the 1990s, however, local politicocultural movements have emphasized the distinctive character of some varieties of Lahnda, and the southern Seraiki (Ser) (Saraiki) and to a lesser extent the northern Hindko (also termed Pahari or Panjistani) are beginning to be employed as independent literary languages consciously rivaling Pakistani Panjabi (Pj). Written in the Perso-Urdu script, all three incipient standards are considerably influenced by Urdu.

Siraiki

The rather homogeneous southwestern Seraiki dialects (Shackle, 1976) are notably distinguished from Panjabi by the retention of historical aspiration and the development of four implosive phonemes /ɟ, ʃ, ɖ, ɓ/, shared with Sindhi, thus Sanskrit *baddha*- ‘bound’ > Pj /bādda/ but Ser /ʔ bāddha/, Sindhi /ʔ bādho/.

Distinctive Seraiki morphological features include the passive extension /-i-/ and the sigmatic future, with stressed extension of transitive stems, e.g., Ser /kəʔresi/ ‘will do’, passive /kəʔrisi/ versus Pj /kəʔrega, ‘kita jaʔvega/; and a full set of suffixed pronouns, often entailing shifted stress, e.g., /ʔkita/ ‘did’,

/ʔkitum/ ‘I did’, /kɪʔtose/ ‘we did’. Common lexical distinctions include Ser /vəʔpaŋ/ ‘to go’, future /vəsi/, versus Pj /jaŋa, jaʔvega/, and the Ser objective and ablative postpositions /kū, kənū/ versus Pj /nū, tō/. Typical Seraiki shibboleths appear in /əʔkhiomis jo sakū bəhū jəldi vəʔŋa posi/ [*told-me-him that us- to very quick going fall-will*] ‘I told him that we should have to go very early’, where /əʔkhiomis/ has a double suffix (/ʔakhia/ + 1sg /-m/ + 3sg /-s/), versus the analytic Pj /mæ oʔnū akhia pəi sanū bəʔt jəldi jaŋa pəʔvega/.

Hindko

Hindko is the term locally used to cover the heterogeneous northern dialects spoken in the hilly areas above the Salt Range (Shackle, 1980). These include the well-described Awankari (Aw) (Bahri, 1963) and the very different Hindko of Peshawar city (Pe). Like Panjabi, Hindko has tonal realizations of historical aspiration, but the phonetic features associated with the Panjabi low-rising tone accompany a high-falling tone in Peshawar Hindko, e.g., Sanskrit *bhāra*- ‘load’ → Aw, Ser /bhar/, versus Pe /pʔār/ (Pj /pʔār/). The Seraiki sigmatic future and pronominal suffixes are shared by Hindko, where the model sentence would appear as Aw /mæ usā akha jo bəʔū jəldi vəʔŋa posi/, Pe /məne unū kiʔa ke bəʔt jəldi jaŋa pəsi/.

Bibliography

- Bahri H (1963). *Lahndi phonetics*. Allahabad, India: Bharati Press.
- Shackle C (1976). *The Siraiki language of central Pakistan: a reference grammar*. London: SOAS.
- Shackle C (1980). ‘Hindko in Kohat and Peshawar.’ *Bulletin of SOAS* 43, 482–510.
- Shackle C (2003). ‘Panjabi.’ In Cardona G (ed.) *The Indo-Aryan languages*. London: Routledge.

Lak

V A Friedman, University of Chicago, Chicago, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Lak language (*Ethnologue* code LBE) belongs to the Daghestanian branch of the Nakh-Daghestanian family and has over 200 000 speakers, mostly in the Republic of Daghestan, Russian Federation (maps of the region are available through the website ‘Thesaurus Indogermanischer Text- und Sprachmaterialien (TITUS)’; see Relevant Websites, at the end of this article). The Lak self-designation is *Lak* (adjective *Lakˀu*); other terms include Turkish *Beyaz lezgi* ‘White Lezgian’ and *Kazikumux*, after Kumux, the main aul (village) of Lakkia and the former center of a feudal state. Lak glosses started around 1600; whole texts appeared in 1734. Lak was written in the Arabic Alphabet until 1928, then in Latin until 1938, and finally in Cyrillic. Lak has five dialects; the standard language is based on Kumux.

Phonology

The Lak vowels are /a, i, u/, all of which can be distinctively pharyngealized, which results in their allophonic centralization [æ^ɣ, e^ɣ, ø^ɣ]. The Lak consonants are shown in Table 1. Geminate (emphatic) consonants are realized as simple unaspirated, except in prevocalic (and, for stops, noninitial) position. Consonant labialization is distinctive in some dialects, and vowel length and stress interact.

Morphology

Lak has four noun classes: I = male sentient, II = mature female sentient, III = other animate and some inanimate, and IV = inanimate and a few lower animals. Any part of speech can take class agreement markers, which are prefixed, infix, and/or suffixed. Lak nouns have four stems, e.g., nominative (= absolutive) singular, as in *q:at:a* ‘house’; oblique singular, as in *q:at-lu-*; nominative plural, as in *q:at-ri-*; and oblique plural, as in *q:at-ra-*. There are more

than 30 stem formants. Case endings attach agglutinatively to the oblique. The three core cases are nominative-Ø, genitive *-l*, and dative *-n*. Secondary cases include addressive/possessive *-x̣*, admotive/dative *-2-x̣:un*, ablative/involuntative/possibilitive *-š:a*, and comitative *-š:al*. A few other affixes attach to the oblique and/or nominative stems.

Lak has six oblique locational affixes (LA) to which five movement affixes (MA) can attach, forming potentially 30 local cases. Not all local cases occur in all combinations or with all substantives, but, unlike spatial postpositions (which take the genitive), local cases have nonlocational functions and cannot govern across a conjunction (abbreviations: IMP, imperative; SG, singular; NOM, nominative; GEN, genitive; EMPH, emphatic; OBL, oblique; PL, plural; INF, infinitive).

| | | |
|----------------------------|-----------------|---------------------------|
| o ^ɣ wč-a | wi-l-a | dus-tura-j-n |
| invite-IMP | you.SG-GEN-EMPH | friend-OBL.PL-ON.LA-TO.MA |
| wa malla-na-j-n | pulaw | b-uk-an |
| and mullah-OBL-ON.LA-TO.MA | pilaf(III) | III-eat-INF |

‘Invite your friends and the mullah to eat pilaf’.

Lak has five sets of deictics, which also serve as third-person markers: *wa*, near the speaker, new information; *mu*, relevant to the addressee, old information; *ta*, opposite, level (older unmarked); *ga*, unmarked (older below speaker); and *k’a*, above speaker. The verb has three aspectual stems, perfective/unmarked (*buc-in* ‘bring-INF’), durative in *-la-* (*buc-la-n*), and iterative in *-awa-* (*buc-awa-n*). Synthetic forms of the marked aspects occur only in the present and future. The infinitive serves as the base for the future. Past tense forms usually have a class marker infixed before the last root consonant, and an infixed negator (indicative *q:a-*, imperative *ma-*) precedes the infixed class marker. The verb has numerous synthetic and analytic paradigms marking aspect, tense, mood, and evidential, some with marking for person as well as class and number.

Syntax

Lak is basically object-verb, attributive/genitive-head; it has pragmatically conditioned free word order and a mixed ergative/accusative system. The converbs *ban* ‘to do’ and *x̣un* ‘to become’ are the most frequent markers of transitive/causative vs. intransitive, respectively. For the agent of an ordinary transitive verb, personal pronouns (first and second person) remain in the nominative; other agents take the genitive, which also functions as ergative. Case assignment and verb agreement also depend on the semantics of the verb, focus, and the pragmatic implications of the clause. Experiencers take dative; ability

Table 1 Lak consonants

| | | | | | | | | | | | | |
|---|---|----|----|---|----|----|---|----|-----|---|---|---|
| b | p | p: | p’ | | | | | | | w | | m |
| d | t | t: | t’ | c | c: | c’ | z | s | s: | r | l | n |
| | | | | č | č: | č’ | ž | š | š: | j | | |
| g | k | k: | k’ | | | | | ṣ | ṣ: | | | |
| | q | q: | q’ | | | | ğ | ṣ̣ | ṣ̣: | | | |
| | | | | | | | | h | | | | |
| | ʔ | | | | | | | h | | | | |

and accident are marked by ablative. Complement clauses trigger class III agreement. The following sentences are illustrative (a resumed morpheme interrupted by a class marker is indicated by <\$>; GER, gerund; PRES, present; ABL, ablative; DAT, dative; ABS, absolutive; DUR, durative aspect; PA, perfective aspect).

ga-na-l k'ili d-a-r-ř:unu
he-OBL(II)-GEN saddle(IV) IV-sell-IV-\$-PAST.GER
Ø-u-r
I-be-3SG.PRES
'He has sold the saddle'.

ga-na-l k'ili d-a-r-ř:unu
he-OBL(II)-GEN saddle(IV) IV-sell-IV-\$-PAST.GER
d-u-r
IV-be-3SG.PRES
'Apparently he sold the saddle'.

ga-na-ř:a k'ili d-a-r-ř:unu
he-OBL(II)-ABL saddle(IV) IV-sell-IV-\$-PAST.GER
d-u-r
IV-be-3SG.PRES
'He accidentally sold the saddle'.

ga-na-n k'ili d-a-ř:an č:a-j
he-OBL(II)-DAT saddle(IV) IV-sell-INF want-PRES.GER
b-u-r
III-be-3SG.PRES
'He wants to sell the saddle'.

ga-na-ř'a k'ili d-a-ř:an
he-OBL(II)-ABL saddle(IV) IV-sell-INF
b-u^čq-la-j b-u-r
III-can-DUR-PRES.GER III-be-3SG.PRES
'He can sell the saddle'.

ni-ti-l q:at-lu-w-un-m-aj
mother(II)-OBL-GEN house-OBL-in(LA)-
toward(MA)-III-\$
č:at' la-w-s-un na-j
bread(III) bring-III-\$-PA.GER go-PRES.GER
b-u-r
III-be-3SG.PRES
'Mother brings bread into the house'.

ninu q:at-lu-wun-m-aj
mother(II).ABS house-OBL-in(LA)-toward(MA)-III-\$
č:at' la-w-s-un na-j
bread(III) bring-III-\$-PA.GER go-PRES.GER
d-u-r
II-be-3SG.PRES
'Bread is brought by mother into the house'.

ninu q:at-lu-wun-n-aj
mother(II).ABS house-OBL-in(LA)-toward(MA)-II-\$
č:at' la-w-s-un na-j d-u-r
bread(III) bring-III-\$-PA.GER go-PRES.GER II-be-3SG.PRES
'It is mother who brings bread into the house'.

Bibliography

- Anderson G (1997). 'Lak phonology.' In Kaye A (ed.) *Phonologies of Asia and Africa* (including the Caucasus). Winona Lake, IN: Eisenbrauns. 973–997.
- Burč'uladze G (1986). *Brunvata Šedgenilobisa da arsebit saxelta brunebis p'rocesebis dziritadi sak'itxebi lak'ur enaši* [Fundamental questions of case composition and processes of declension of the noun in the Lak language]. Tbilisi: Mecniereba.
- Burč'uladze G (1987). *Lak'uri zmna* [The Lak verb]. Tbilisi: Mecniereba.
- Friedman V (1992). 'Lak substantival declension: 40 cases or 50?' In Aronson H (ed.) *The non-Slavic languages of the USSR: linguistic studies – second series*. Chicago: Chicago Linguistic Society. 113–133.
- Friedman V (1994). 'Ga in Lak and the three "there"s: deixis and markedness in Daghestan.' In Aronson H (ed.) *NSL.7: linguistic studies in the non-Slavic languages of the commonwealth of independent states and the Baltic republics*. Chicago: Chicago Linguistic Society. 79–93.
- Friedman V (2004). 'The expression of speaker subjectivity in Lak.' In Guentchéva Z (ed.) *L'Énonciation médiatisée 2*. Louvain: Peeters. 75–83.
- Helmbrecht J (1996). 'The syntax of personal agreement in East-Caucasian languages.' *Sprachtypologie und Universalienforschung* 49, 127–148.
- Kazenin K (1998). 'On patient demotion in Lak.' In Kulikov L & Vater H (eds.) *Typology of verbal categories*. Tübingen: Max Niemeyer. 95–115.
- Kazenin K (2001). 'Focus in Daghestanian languages and word order parameter.' *Linguistic Typology* 1, 289–316.
- Khajdakov S (1966). 'The dialect-divisions of Lak.' *Studia Caucasica* 2, 9–18.
- Khalilov Kh (1976). *Jumor I satira narodov Dagestana* [Humor and satire of the peoples of Daghestan]. Max-ačkala: AN SSSR.
- Kibrik A E (1997). 'Beyond subject and object: toward a comprehensive relational typology.' *Journal of Linguistic Typology* 1, 279–346.
- Murkelinskij G (1971). *Grammatika laskogo jazyka*. [Grammar of the Lak language] Maxačkala: Dagučpedgiz.
- Žirkov L (1955). *Lakskij jazyk* [The Lak language]. Moscow: AN SSSR.

Relevant Websites

<http://titus.fkidg1.uni-frankfurt.de> – The website of The-saurus Indogermanischer Text- und Sprachmaterialien (TITUS). Maps of Asia and the Caucasus region are available by selecting the links 'Teaching Materials,' then 'Language Maps'.

Lakota

B Ingham, SOAS, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Lakota is one of a group of closely related dialects sometimes referred to by linguists as Dakotan. These include Lakota in the west, Dakota in the east, Nakota in the north, and Nakoda in the northwest. The speakers of Lakota and Dakota were traditionally referred to in English as the Sioux, those of Nakota as the Assiniboiné, and those of Nakoda as the Stoney. Lakota and Dakota are mutually intelligible. Reports differ as to how far Nakota and Nakoda are intelligible with the other two. Dakotan is part of a group of languages known as Siouan-Caddoan centered mainly on the central plains and prairies, but also represented on the eastern seaboard.

Reports on the number of speakers of Lakota range from 6000 to 20 000. Great efforts are being made to preserve the language in schools, colleges, and universities in the region and there is probably a considerable degree of partial or receptive knowledge of it.

Morphology

Major word classes of Lakota are noun, verb, adverb, postposition, demonstrative, pronoun, and conjunction. The verb in particular can be regarded as polysynthetic and noun incorporation occurs in the verb and adverb. The functions often covered by adjectives in other languages are covered in Lakota by stative verbs and adverbs.

The verb system is of the split intransitive type, where agents occur only in the Active verb class while patients occur with the Active and Stative types. These are marked in the verb by prefixes or infixes as shown below:

| | |
|--------------------------|---------------------------|
| agent marker | patient marker |
| wa-‘I’ | ma-‘me’ |
| ya-‘you’ | ni-‘you’ |
| uŋ(k)-‘we’ (‘you and I’) | uŋ(k)-‘us’ (‘you and me’) |
| | wic’a-‘them’ (animate) |

composite pronoun prefix
c’i-‘I (agent)-you (patient)’

Plurality in the third and second person and inclusiveness of third persons in the first plural is marked by a suffix *-pi*. The occurrence of these markers with the two verb types is shown below:

| | |
|--------------------------------|-------------------------|
| Active verb <i>u</i> ‘to come’ | |
| sing, exclusive | plur, inclusive |
| wau ‘I come’ | |
| yau ‘you (sing) come’ | yaupi ‘you (plur) come’ |

| | |
|------------------------------|----------------------------------|
| <i>u</i> ‘he, she, it comes’ | <i>upi</i> ‘they (animate) come’ |
| <i>uŋku</i> ‘we (excl) come’ | <i>uŋkupi</i> ‘we (incl) come’ |

| | |
|-------------------------------------|---------------------------------------|
| Stative verb <i>k’uja</i> ‘be ill’ | |
| sing, exclusive | plur, inclusive |
| <i>mak’uje</i> ‘I am ill’ | |
| <i>nik’uje</i> ‘you (sing) are ill’ | <i>nik’ujapi</i> ‘you (plur) are ill’ |
| <i>k’uje</i> ‘he is ill’ | <i>k’ujapi</i> ‘they are ill’ |
| <i>uŋk’uje</i> ‘we (excl) are ill’ | <i>uŋk’ujapi</i> ‘we (incl) are ill’ |

Forms with both agent and patient prefixes, showing some infixed, are shown with the active verb *ole* ‘to seek’: *owic’awale* ‘I seek them’, *owic’ayale* ‘you (sing) seek them’, *oc’ile* ‘I seek you (sing)’, *oc’ilepi* ‘I seek you (plur)’, *onile* ‘he seeks you (sing)’, *uŋkonilepi* ‘we (incl) seek you’, *omayale* ‘you (sing) seek me’, *uŋkoyalepi* ‘you (plur) seek us (incl)’.

Nouns do not show number distinctions, but may mark a possessor as in *mi-hingna* ‘my husband’, *ni-hingna* ‘your husband’, *mi-c’iŋksi* ‘my son’, *c’iŋksitku* ‘his, her son’. Other nominal prefixes may distinguish noun classes such as instruments as in *wi-c’ap’e* ‘fork’, abstract concepts as in *wo-slolye* ‘knowledge’, time and space concepts as in *o-ap’e* ‘hour’, *o-mak’a* ‘year’, *o-nap’e* ‘refuge’.

Word Derivation

Word derivation is by affixation and compounding. Adverbs are formed with a suffix *-ya*, *-yela* or *-yaŋ* from stative verbs as in *wašte* ‘be good’, *wašte-ya* ‘well’, *ska* ‘be white’, *ska-yela* ‘whitely’, *wak’aŋ* ‘be sacred’, *wak’aŋ-yaŋ* ‘in a sacred manner’. Adverbs are widely used to state characteristics of objects often with the verb *haŋ/he* ‘stand’ as in *małhiya waŋ ska-yela he* ‘a cloud stood whitely’ meaning ‘there was a white cloud’, *op’osya he* ‘it stood coldly and clearly’ meaning ‘the weather was cold and clear’. Postpositions can be formed from adverbs by a prefix *i-* as in *wiyohpeyata* ‘in the west’, *i-wiyohpeyata* ‘to the west of’.

A set of circumstantial stems are important in forming verbs, adverbs, wh-words and postpositions as in *tok’eca* ‘be somehow’, *hec’eca* ‘be like this’, *iyec’eca* ‘be like’, *tok’el* ‘how’, *hec’el* ‘like this’, *iyec’el* ‘like, as’ from a stem **-k’ec’a/-k’el* indicating ‘quality’. Other such stems are **-haŋ* ‘time’, **-k’etu* ‘occurrence’, **-nakeca* ‘number’, **-haŋkeca* ‘(extent’, yielding among others *hehaŋ* ‘then’, *tohaŋ* ‘when’, *iyehaŋtu* ‘be time for’, *hec’etu* ‘happen thus’, *iyec’etu* ‘happen as’, *iyenakeca* ‘be as many as’, *iyena* ‘as many as’, *tonakeca* ‘how many’, *tohaŋyaŋ* ‘how far’, *iyehaŋyaŋ* ‘as far as’.

Syntax

Word order is generally agent-patient-verb as in *wic'asa ki c'ingksitku ki ole* 'the man looked for his son.' Relative clauses are marked by the use of *wan* 'one', *ki* 'the' and *he* 'that' as in *wic'asa wan c'ingksitku ki ole ki he owale* 'I looked for the man who was looking for his son.' Sentences can be embedded in higher sentences by using *ki/k'uŋ* 'the' and by certain postpositions as in *c'ingksitku ki ole ki slolwaye* 'I know that he was looking for his son', *c'ingksitku t'i ekta wau* 'I came to where his son lived'.

Nouns can be incorporated into verbs and verbs subordinated to other verbs by preposing, often also with stem truncation, as in *šung-ole* 'he looked for horses', *šung-manuŋ* 'he stole horses', *šung-ole-mani* 'he traveled looking for horses' (*šung-* < *šungka* 'horse, dog'), *iŋyang-mani* 'he traveled running' (*iŋyang-* < *iŋyangka* 'run'), *kali-si* 'ask to make' (*kali-* < *kaga* 'make').

Men's and Women's Speech

These are distinguished by certain sentence-final particles of high frequency of occurrence shown below.

| | male speaker | female speaker |
|-------------|---------------------|-----------------|
| declarative | -yelo, -welo, -ksto | -ye, -we, -ksto |

| | | | |
|---------------|------|-----------------------|----------|
| interrogative | | -hwo (-he informally) | -he |
| imperative | sing | -yo, -wo | -ye, -we |
| | plur | -po | -pe |

Bibliography

- Boas F & Deloria E C (1941). *Dakota grammar*. Washington: U. S. Government Printing Office.
- Buechel E (1939). *A grammar of Lakota: the language of the Teton Sioux Indians*. Rosebud SD: Rosebud Educational Society.
- Buechel E (1970). *Lakota English dictionary*. Pine Ridge SD: Red Cloud Indian School.
- Buechel E (1978). *Lakota tales and texts*. Pine Ridge SD: Red Cloud Indian School.
- Bushotter G (1887). *Bushotter papers: texts by George Bushotter 1887*. Manuscript 4800 Dorsey Papers: Dakota: Teton (3.1.1.3) 103. Smithsonian Institution National Anthropological Archives.
- Deloria E C (1932). *Dakota texts*. Publications of the American Ethnology Society. Boas F (ed.). New York: G. E. Stechert. Reprinted in 1974. New York: AMS Press.
- Ingham B (2001). *English-Lakota dictionary*. Richmond, Surrey: Curzon Press.
- Ingham B (2003). *Lakota: languages of the world/materials* 426. Munich: Lincom Europa.
- Parks D R & DeMallie R J (1992). 'Sioux, Assiniboiné and Stoney dialects: a classification.' *Anthropological Linguistics* 34(i-iv), 233–256.
- Riggs S R (1893). *Dakota grammar texts and ethnography*. Washington: Government Printing office.
- Rood D S & Taylor A R (1976). *Lakhota grammar or beginning Lakhota* (vol. 2). Boulder: University of Colorado Lakhota Project.

Lao

D A Smyth, SOAS, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Number of Speakers and Genetic Relationship

Lao belongs to the southwestern subgroup of the Tai language family. It is the national language of Laos and the first language of approximately 70% of the population of 5.4 million. There is considerable regional variation in dialects but the Vientiane dialect is regarded as standard Lao and serves as a lingua franca for the country's many ethnic minorities. Lao dialects are spoken by a further 12–15 million people in northeast Thailand, and there are sizeable overseas Lao communities in both France and the United States.

Lao bears many very close similarities to Thai and, because of Thailand's economic and cultural dominance, most Lao people understand spoken Thai. Thai radio and television broadcasts, Thai pop songs, videos of Thai movies, and foreign movies dubbed in Thai are all popular in Laos; many educated Lao people also can read Thai. Thais, however, have more difficulty understanding Lao, partly through limited exposure to the language, and partly through a lack of desire to understand it, as they regard Lao as a language of low prestige.

Phonology and Grammar

Lao is a tonal language. Standard Lao has six tones: low, mid, high, rising, high falling, and low-falling. The vowel and consonant systems closely resemble

those of standard Thai; notable difference are the existence of an initial /p/ and the absence of /r/ in spoken Lao. Grammatically, too, there are close similarities to Thai: word order is *subject-verb-object*, nouns and verbs are not inflected, the pronominal system is complex and capable of conveying subtle degrees of relative status and intimacy. 'Classifiers' or 'count words' are used in noun phrases involving numbers.

Words of purely Lao origin are often monosyllabic. Sanskrit and Pali borrowings are numerous, and where they coexist with an indigenous Lao word they reflect a more formal or literary style. Other sources of loan words are Thai, Chinese, and Cambodian, although with Thai and Lao sharing many common basic words, the extent of Lao borrowing can be overestimated; many relatively recently coined Thai words have, however, been consciously absorbed into Lao. Despite the country's former colonial status, French loan words are relatively few.

Sample of Lao with Translation

| | | | |
|-------------------------------|-----------------|-----------------|------|
| khôy | dây | hùucák | káp |
| 1st pers. pron. | to get to | to know (s.one) | with |
| láaw | jūu | hóonhian | |
| 3rd pers. pron. | location marker | school | |
| 'I got to know her at school' | | | |

Recent History

When the boundaries of present-day Laos were drawn up in 1893 under the terms of a Franco-Siamese treaty, the Lao-speaking population was divided in two, the majority paradoxically being in

northeast Thailand. The French brought in Vietnamese to carry out much of the administration of their colony, and with French the medium for what little postprimary education existed, the Lao language suffered a loss of prestige, even among many of its own speakers. The decline of French influence and the rise of nationalism in the aftermath of World War II helped to improve the status of Lao. Although the communist government, which came to power in 1975, has Lao-ized the education system, introduced adult literacy programmes and attempted to teach Lao to the country's ethnic minorities, literacy rates remain low.

Bibliography

- Chamberlain J R (1978). 'Language standardisation in Laos.' In Perez A Q, Santiago A O & Nguyen Dang Liem (eds.) *Papers from the conference on the standardisation of Asian languages, Manila, Philippines, December 16–21, 1974*. Pacific Linguistics, Series C – no. 47. Canberra: Research School of Pacific Studies, Australian National University.
- Hoshino T & Marcus R (1981). *Lao for beginners*. Rutland, Vermont. Charles E. Tuttle.
- Enfield N J (1999). 'Lao as a national language.' In Evans G (ed.) *Laos: Culture and society*. Chiang Mai. Thailand: Silkwork Books. 258–290.
- Morev L N, Moskalyov A A & Plam Y Y (1979). *The Lao language*. Moscow: USSR Academy of Sciences, Institute of Oriental Studies.
- Smalley W A (1994). *Linguistic diversity and national unity: language ecology in Thailand*. Chicago: University of Chicago Press.
- Yates W G & Souksomboun S (1970). *Lao basic course*. Washington, D.C.: Foreign Service Institute.

Latin

J Powell, Royal Holloway, University of London, Egham, Surrey, UK

© 2006 Elsevier Ltd. All rights reserved.

History and Affiliations

Latin is an Indo-European language of western central Italy, recorded from about 500 B.C. in the immediate area of Rome, related closely to the poorly attested Faliscan dialect and more distantly to Osco-Umbrian and Venetic. Lexical influences from

neighboring languages (Italic dialects, the unrelated Etruscan language, and above all Greek) are visible in Latin from the earliest period. With the rise of Rome, Latin spread throughout Italy and into the provinces of the Roman Empire. It became the spoken language in western continental Europe, displacing the indigenous languages; its spoken varieties developed into the modern Romance languages. A more or less standard version of written Latin remained in use in administration, law, education, and the Church, and in due course spread throughout Europe as Medieval Latin. The status of Latin as a common medium of

learned communication began to wane only with the rise of English as a world language in modern times.

Varieties

The dialect of Rome became the dominant form of Latin from an early period, with variations from it being regarded as rustic. Written Latin was standardized by the first century B.C. and remained broadly uniform thereafter, concealing changes and diversification in the spoken idiom; nevertheless a range of literary, technical, or colloquial registers can be distinguished in the surviving texts, and the occurrence of nonstandard written forms can provide evidence for spoken developments. The term *Vulgar Latin* is used to refer to any variety that departs significantly from classical norms. There is no good ground for regarding later written forms – Christian, Medieval, Renaissance, or Modern Latin – as essentially different varieties of the language: despite differences, often quite wide, in orthography, lexicon, style, and pronunciation, they all reflect the same basic linguistic structure and a common if diverse literary tradition. The following approximate periods may be recognized for convenience: Early Latin, to the 2nd century B.C.; Classical Latin, divided into Republican, and Augustan ('Golden'), and Early Imperial ('Silver'), until the 2nd century A.D.; Late Latin (i.e., the Latin of Late Antiquity), 3rd to 6th century A.D.; Medieval Latin, from the 7th to the 13th century A.D., including two high points of Latin culture, the Carolingian period (9th century) and the 12th-century renaissance. The term *Neo-Latin* embraces Renaissance Latin (14th to 15th centuries), the scholarly and scientific Latin of the early modern period, and modern uses of the language.

Script

The Latin alphabet, the basis of current Western European alphabets, was derived from a form of Greek script introduced into Italy by colonists from Euboea. In its classical form, it contains all the letters that are still used in Modern English except for J, U, and W; V (the U shape developed later) was used for the vowel /u/ and the semivowel /w/; I was used for the vowel /i/ and the semivowel /j/. Y and Z, originally absent from Latin, were imported for use in Greek borrowings. The capital letters, used for inscriptions, have retained their classical form more or less unchanged. Originating in Roman cursive, a variety of handwritten styles evolved from late antiquity onward and ultimately gave rise to modern lowercase letters.

Phonology

Sound/Spelling Relationship

The Classical Latin alphabet represents the sounds of the language quite well, except for its failure to distinguish the semivowels (as we now distinguish, e.g., *voluit* 'he wanted' from *volvit* 'he rolls'). Phonemic doubling of consonants was usually but not invariably indicated. In the classical period, vowel length was shown sometimes by doubling the vowel, more often by means of the *apex*, a diacritic resembling the acute accent. In Latin as conventionally printed, vowel length is not regularly indicated; so, for example, *canis* can represent either *cānīs* 'dog' or *cānīs* 'grey hair (dative-ablative)'.

A number of sounds initially caused spelling difficulties: (a) Classical Latin had a series of nasal vowels, represented (consistently with etymology) by vowel + *m* finally and vowel + *n* medially; (b) a vowel sound midway between /i/ and /u/ was found in words such as *optumus/-imus* 'best' and other superlatives, and *lubens/libens* 'willing'; (c) there was no separate symbol for the velar nasal (see note *c* under the following consonant table).

Consonants

The consonantal inventory of Latin may be set out as follows:

| | Labio-velar | Velar | Dental | Labial |
|----------------|-------------------------------------|--|--------|----------------------------------|
| Voiceless stop | /k ^w / <qu> | /k/ ^a <c>/ <k> | /t/ | /p/ |
| Voiced stop | /g ^w / <gu> ^b | /g/ | /d/ | /b/ |
| Nasal | | /ŋ/ <n>/ <g> ^c | /n/ | /m/ |
| Fricative | glottal: /h/ | alveolar: /s/ | | labiodental: /f/ |
| Liquid | | rolled/ flapped: /r/ lateral: /l/ ^d | | |
| Semivowel | | palatal: /j/ <i> | | labial: /w/ <u>/<v> ^e |

^aThe original sound was /k/, but the Romance and Medieval palatal realization before front vowels may already have started to develop in antiquity.

^bOccurs only medially after a nasal or liquid as in *languidus* 'languid', *urgueo* 'press'.

^cOccurs only medially, (a) as the outcome of /n/ before a velar, or (b) as the outcome of /c/ or /g/ before a nasal; written in the first case as <n>, as in *tango* 'touch', and in the second as <g> as in *dignus* 'worthy' (<*dec-nos, cf. *dec-ens* 'decent'). The combination /ŋn/ <gn> later underwent a further sound change to palatal /ɲ/, found, for example, in Italian.

^dRepresented by both 'clear' and 'dark' (velarized) allophones, as shown by the effect on neighboring vowels.

^eThe sound /w/ later changed to /β/ or /v/; spelling variations indicate that this took place (under Greek influence?) in Italy and the East possibly as early as the first century A.D., and in the western provinces at a later period.

Vowels

Classical Latin had a symmetrical five-way vowel system with phonemic distinction between long, short, and nasalized vowels: e.g., *puellā* ‘girl’ (nominative): *puellā* /-a:/ (ablative): *puellam* /-ā/ (accusative). Early Latin had a range of inherited diphthongs, which were considerably simplified by the Classical period; further leveling took place in varieties of the spoken language (*au* > /o:/, *ae* > open /ɛ:/, *oe* > /e:/), giving rise to confusion in spelling. In Romance, the classical vowel system was remodeled: nasalization and phonemic vowel length disappeared, while changes in vowel quality led to merger of short /i/ with long /e:/, and of short /u/ with long /o:/.

Accent and Prosody

Classical and post-Classical Latin has a largely predictable stress accent: the accent falls on a long penultimate, but on the antepenultimate if the penultimate is short. (A short syllable is an open syllable containing a short vowel; all other syllables, except for a few doubtful cases, are long.) This had apparently replaced an earlier initial stress accent, which is postulated to account for the regular vowel weakening that took place in early Latin in non-initial syllables, e.g., *pro* ‘before’ + *habeo* ‘hold’ > *prohibeo* ‘forbid’. The usually clear distinctions of syllabic quantity made it possible for quantitative meters borrowed from Greek to replace the native ‘Saturnian’ verse form. Stress accent plays no formal part in Classical Latin verse, although there is an interaction between quantitative rhythm and word accent, which both creates some metrical restrictions and allows for considerable poetic subtlety. Stressed rhythms were reintroduced in Late and Medieval Latin.

Morphology

Inflection

Latin is an inflected language retaining a number of features of the parent Indo-European, but with considerable innovation, especially in the verbal paradigm.

Nouns are inflected for number and case. There are two numbers, singular and plural, and five regular cases: nominative, accusative, genitive, dative, and ablative. Nouns are traditionally grouped in five declensions: the first and second declensions continue the IE O/A stems, the third and fourth continue the IE stems in consonants and semivowels, and the fifth is a Latin innovation of mixed origin. A separate vocative is found only in the second declension and in borrowed Greek names. A locative case exists in place names and in a few common nouns, such as *domi*

‘at home’ from *domus* ‘home’. Nouns are assigned to one of three gender categories – masculine, feminine, and neuter – on the basis partly of sense and partly of form, so that on the whole there are fewer unpredictable or illogical assignments of gender than in many other languages.

Adjectives, both as attributes and as complements, agree with nouns in gender, number, and case; they broadly follow the same patterns of declension as nouns. There are no articles, but there is a rich pronominal system including four demonstratives (*is* ‘the one mentioned’, *hic* ‘this’, *iste* ‘that of yours’, *ille* ‘that over there’) and a distinction between specified and nonspecified indefinite pronouns (*quidam* ‘someone in particular’, *aliquis* ‘someone or other’).

Verbs are inflected according to the person and number of the subject. Subject pronouns are regularly dropped (*amamus* ‘we love’). There are two sets of personal endings, indicating voice: active -*o/-m -s -t -mus -tis -nt*, passive -*r -ris/-re -tur -mur -mini -ntur*; and one other set of endings peculiar to the perfect active. Verbs are classified in four conjugations according to stem vowel. The principal parts of a Latin verb, from which all other forms can be derived, are the first person singular present indicative active (normally listed in the dictionary, e.g., *dico* ‘I say’), the present infinitive active (e.g., *dicere* ‘to say’), the first person singular perfect active (e.g., *dixi* ‘I said’), and the past participle passive (e.g., *dictum* ‘said’). The stems of the present and perfect are combined transparently with tense markers to form two further indicative tenses and two subjunctive from each stem, e.g., *procedeba-m* ‘I was proceeding’ (imperfect indicative), *praedix-er-a-tis* ‘you had predicted’ (pluperfect indicative), *mane-re-mus* ‘we would be staying’ (imperfect subjunctive), *decipi-a-mini* ‘you would be deceived’ (present subjunctive passive). There are two imperatives, one used for immediate commands (*exi!* ‘get out!’), the other for instructions (*bene misceto* ‘mix well’). The perfect tenses of the passive are formed periphrastically: the past participle is combined with an appropriate (present-stem) form of *esse* ‘be’, e.g., *occisa est* ‘she has been killed’.

A restricted class of verbs (generally known as ‘deponents’: fossilized remnants of the old middle voice) take passive morphology although their meaning is active and can be transitive: e.g., *philosophor* ‘I philosophize’; *vereor* (transitive) ‘I fear’. Intransitive active verbs (including a class of verbs that take a dative object) have no passive, except third-person singular forms used impersonally, e.g., *venio* ‘I arrive’; *venitur* ‘people arrive’ (cf. the impersonal passive in the Celtic languages). Genuinely irregular verbs are few (*sum* ‘be’, *possum* ‘can’,

volo ‘want’, and some others). The greatest difficulty in Latin verbal morphology is the often unpredictable formation of the perfect stem.

The verb has two infinitives, present (*dicere* ‘to say’) and perfect (*dixisse* ‘to have said’); other verbal nouns (gerunds, supines) supply case forms for the infinitive. There are three participles, present active (*dicens* ‘saying’), past passive (*dictum* ‘said’), and future active (*dicturus* ‘about to say’), and a verbal adjective (the gerundive), which is future passive in meaning and often denotes obligation (*faciendum est* ‘it is to be done’).

Derivational Morphology

Latin has a productive system of derivational morphology, much of which is familiar through derivatives in modern languages, e.g., prepositional prefixes such as *ex-* ‘out’, *circum-* ‘round’; other prefixes such as *in-* ‘not’, *re-* ‘back’, nominal suffixes such as *-tio(n)-* (process) and *-tor* and *-trix* (male and female agents), and adjectival suffixes such as *-bilis* (*inextricabilis* ‘inextricable’), *-anus* (*Christianus* ‘Christian’), *-arius* (*piscarius* ‘to do with fish’). Adjectives form adverbs (in *-e*, *-o*, or *-ter*), comparatives in *-ior*, and superlatives in *-issimus*. At first, Latin was resistant to some kinds of derivation; for example, compound nouns or adjectives were relatively unusual, occurring mostly in high style (*solivagus* ‘wandering alone’) or as comic colloquialisms (*caldicerebrius* ‘hot-headed’). From the classical period onward, the need to express new ideas led to a large expansion of the classical lexicon both by derivation and by borrowing.

Syntax

Word Order

Because Latin encodes basic grammatical relations by means of inflection, there is considerable freedom in word order. Either a topic or an emphasized predicate may be placed first in a sentence, and variations of logical emphasis are also common at phrase level. Classical Latin preserves the Indo-European phenomenon known as Wackernagel’s Law, i.e., the first stressed position in a sentence or clause is followed by an unstressed position which often contains, in order of precedence, (a) sentence connectives, (b) weak pronouns, (c) unstressed verbs. The phenomenon of hyperbaton, or discontinuity in the noun phrase, is a particular feature of Latin as it is of Classical Greek: its effect is often to throw particular focus on to the first element of the noun phrase, e.g.:

| | | |
|------------------------|----------------|-----------------|
| <i>bonos</i> | <i>habemus</i> | <i>consules</i> |
| good-MASC.ACC.PL. | we-have | consuls-ACC.PL. |
| + Focus | | |
| ‘we have good consuls’ | | |

Because of inflection, the absence of articles, and the flexibility of word order, Latin is capable of considerable compression, hence the lapidary style favored in inscriptions, proverbs, and epigrams.

Complex Sentence Structure

Subordination is much favored in varieties of literary Latin; Classical Roman writers influenced by the Greek periodic style made full use of the available syntactical resources, e.g., the inflected relative pronoun enabling any nominal constituent to be relativized, and the free use of the accusative-infinitive construction to create continuous indirect speech (*oratio obliqua*), as well as paratactic devices such as apposition, parallelism, ellipsis, and chiasmus. Means were found to compensate for the lack of certain features (e.g., the past participle active) which had made for flexibility in Greek, and the rich tense system is fully exploited in classical narrative prose. Late and Medieval Latin usage often blurs the distinctions of tense and mood found in Classical Latin; the conjunction *quod* ‘that’ gains ground at the expense of the accusative-infinitive, the active present participle becomes more frequent, and the periphrastic constructions characteristic of the Romance verbal conjugation begin to appear.

Bibliography

- Adams J N (2003). *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.
- Allen W S (1978). *Vox Latina. The pronunciation of Classical Latin* (2nd edn.). Cambridge: Cambridge University Press.
- Baldi P (1999). *The foundations of Latin*. (Trends in Linguistics: Studies and Monographs 117.) Berlin and New York: Mouton de Gruyter.
- Courtney E (1999). *Archaic Latin prose*. (American Philological Association: American Classical Studies, 42.) Atlanta: Scholars Press.
- Cupaiuolo F (1993). *Bibliografia della lingua latina 1949–1991*. Naples: Loffredo.
- De Meo C (1986). *Lingue tecniche del latino* (2nd edn.). Bologna: Pàtron.
- Ernout A (1953). *Morphologie historique du latin* (3rd edn.). Paris: Klincksieck.
- Ernout A (1954). *Aspects du vocabulaire latin*. Paris: Klincksieck.
- Glare P G W (ed.) (1982). *Oxford Latin dictionary*. Oxford: Oxford University Press.

- Hammond M (1976). *Latin: a historical and linguistic handbook*. Cambridge, MA, and London: Harvard University Press.
- Herman J (1997). *Vulgar Latin*. Wright R (English trans.). University Park, PA: The Pennsylvania State University Press.
- Hofmann J B (1951). *Lateinische Umgangssprache* (3rd edn.). Heidelberg: C. Winter.
- Kennedy B H (1962). *The Revised Latin primer*. Ed. and rev. Mountford, J. London: Longmans.
- Keppie L (2001, 1991). *Understanding Roman inscriptions*. London: Routledge.
- Löfstedt E (1959). *Late Latin*. Oslo: Aschehoug; Cambridge, MA: Harvard University Press.
- Marouzeau J (1953). *L'Ordre des mots en latin* (new edn.) (*Collection d'études latines, série scientifique*) (3 vols). Paris: Les Belles Lettres.
- Marouzeau J (1962). *Traité de stylistique latine* (4th edn.). Paris: Les Belles Lettres.
- Meillet A (1977). *Esquisse d'une histoire de la langue latine*. (3rd edn., rev. Perrot J). Paris: Klincksieck.
- Meiser G (1998). *Historische Laut- und Formenlehre der lateinischen Sprache*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Mohrmann C (1961–1965). *Études sur le latin des chrétiens* (3 vols). Rome: Edizioni di Storia e Letteratura.
- Niedermann M (1953). *Historische Lautlehre des Lateinischen* (3rd edn.). Heidelberg: C. Winter.
- Norberg D (1968). *Manuel pratique de Latin Médiéval*. Paris: A. & J. Picard.
- Palmer L R (1968). *The Latin language*. London: Faber and Faber.
- Pinkster H (1990). *Latin syntax and semantics*. London: Routledge.
- Poccetti P, Poli D & Santini C (1999). *Una storia della lingua latina. Formazione, uso, comunicazione*. Rome: Carocci.
- Raven D S (1965). *Latin metre: An introduction*. London: Faber and Faber.
- Rosén H (1999). *Latine loqui. Trends and directions in the crystallisation of Classical Latin*. Munich: Wilhelm Fink Verlag.
- Sandys J E (ed.) (1921). *A companion to Latin studies* (3rd edn.). Cambridge: Cambridge University Press.
- Stolz F, Debrunner A & Schmid W P (1993). *Storia della lingua latina*. Italian tr. Benedikt C, intr. and notes Traina A (4th edn., rev. Vineis E). Bologna: Pàtron.
- Woodcock E (1959). *A new Latin syntax*. London: Methuen.

Latvian

S Young, University of Maryland Baltimore County, Baltimore, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Latvian (*latviešu valoda*), with some 1.5 million speakers in the Republic of Latvia, is one of two present-day Indo-European Baltic languages, the other being the structurally more conservative Lithuanian. The Latvian standard language is based on the central Latvian dialect (*vidus dialekts*), which is further divided (from west to east) into Couronian, Zemgalian, and Vidzeme varieties. The central dialect, together with Tamian in northwestern Couronia and along the northeastern coast of the Bay of Riga, is known as Low Latvian. High Latvian (Selonian and Latgalian, the latter with an independent literary tradition), is found in the eastern third of the country.

The traditional view of the origin of Latvian is that it represents a synthesis of the language of the early Latgalians (known from the 13th century on simply as *Letti*, that is, Letts or Latvians) and neighboring closely related Baltic dialects, now extinct, among them Zemgalian (along the Lielupe river), Couronian (in southwestern Latvia), and Selonian (along the middle Daugava). The influence of these Baltic substrata, together with that of Balto-Finnic and Slavic

neighbors, accounts in large part for the marked dialectal diversity of Latvian within a relatively small territory.

The beginnings of a Latvian literary tradition date to the Reformation. The first published book in Latvian is a 1585 translation of a Catholic catechism; this was soon followed by other religious texts, chiefly translations. The language of these early texts is to varying degrees influenced by the German speech of the authors, especially in syntax. Efforts to establish a national standard language were begun in the mid-19th century; the Latvian traditional folksongs (*dainas*) served as an important source of norms for the standard language. A milestone in the codification and description of modern Latvian was the appearance of Jan Endzelin's (Jānis Endzelīns) *Lettische grammatik* (1922) and the four-volume *Latviešu valodas vārdnīca* [*Dictionary of Latvian*] (1923–32), begun by Karl Mühlenbach (Kārlis Mīlenbahs) and completed by Jan Endzelin.

The Latvian vowel phonemes include /i, u, ε, æ, a/ (spelled *i, u, e, e, a*) and their long counterparts /i:/, u:/, ε:/, æ:/, a:/ (spelled *ī, ū, ē, ē, ā*). In addition, a long and short *o* are found in borrowed words. The vowels /e/ and /æ/ (and /ε:/ and /æ:/) are represented in writing by a single *e* (*ē*), reflecting the fact that they are derived from a single source, Baltic *e* (**ē*). Originally

conditioned by the nature of the following vowel, they now function as separate phonemes. Functioning as long vowels are the diphthongs [ie, uə] (written *ie*, *o*), which result from Common Baltic **ei* and **ō*: *dīevs* ‘God’ (**deiū-*), *nò* ‘from’ (**nō*), and from tautosyllabic sequences *en* and *an*: *pieci* [pietsi] ‘five,’ *ròka* [rùeka] ‘hand’ (Li. *penkì*, *rankà*). In final syllables Baltic long vowels are shortened and short *e* and *a* are lost: La. *vīlks* ‘wolf’: Li. *vīlks*.

Among the consonants of Latvian, the obstruents reflect a voiced:voiceless opposition, with regressive voicing assimilation (*atbilde* [ˈadɪlɪdɐ] ‘answer,’ *labs* [laps] ‘good’); unlike Lithuanian, word-final consonants in final position are not devoiced: *kad* [kad] ‘when.’ Latvian lacks the feature of palatalization found in Lithuanian, but distinguishes a series of palatal consonants: *k̟* [c], *g̟* [ɟ], *ņ* [ɲ], *l̟* [ʎ]; *k̟* and *g̟* are found mainly in borrowed words. Characteristic of the Latvian consonant system is the development of *ts*, *dz* from Baltic *k*, *g* before a front vowel: *cits* ‘other’ (Li. *kītas*), *dzimt* ‘to be born’ (Li. *gimti*). The Baltic reflexes of the IE palatovelars have merged in Latvian with *s* and *z*: *simts* ‘hundred’ (Li. *šimtas*) < IE **k̑ntó-*, *zeme* ‘earth’ (Li. *žemė*) < IE **d̑h̑em-*. Secondary palatals *f*, *ʒ* (orthographically *š*, *ž*) have arisen in turn from the sequences **s̑i*, **t̑i* (> *f*) and **z̑i*, **d̑i* (> *ʒ*) before a back vowel: *šūt* ‘to sew’ (< **s̑ūt*).

With a few exceptions, Latvian has initial stress. The standard language distinguishes three phonemic tones (marked only in linguistic texts) on long vowels and diphthongs: level [˘] and broken [˙], both representing Baltic acute tone; and falling [ˑ], representing Baltic circumflex: *mīt* ‘to tread,’ *mīt* ‘to change (arch.),’ *mīt* ‘resides.’ Only the central Latvian Vidzeme dialect around Valmiera still distinguishes all three such tones; remaining dialects oppose only two (outside of High Latvian, broken and falling tone typically merge as either falling or broken tone).

The Latvian noun distinguishes masculine and feminine gender, each with three declensional patterns. Within declensional paradigms, five cases are distinguished morphologically: nominative, genitive, dative, accusative, and locative, in both the singular and plural. Adjectives, which occur in both definite and indefinite forms, agree with the noun in number, gender, and case.

The Latvian verb distinguishes present, past, and future tenses, and has a system of relative tenses

formed with *būt* ‘to be’ and the past active participle. The verb has three conjugational types (only two of which are productive), each with a number of subtypes. The 1 sg. present and past forms coalesce in many instances where the stems are not distinguished by root vowel gradation. Like Lithuanian, singular and plural are not distinguished morphologically in the third person.

Peculiar to Latvian, from a Baltic perspective, is the syntactic construction *man ir* ‘to-me is’ for ‘I have’ (Li. *aš turiu*) and the use of the debitive (expressing obligation), formed with the particle *jā-* and the 3rd person present, with a dative subject: *man jābrauc* (‘to-me must-go’) ‘I must go.’ Like Lithuanian, Latvian uses a preposed adnominal genitive: *latviešu valoda* (‘of-Latvians language’ = ‘Latvian language’).

The lexicon has been rather strongly influenced by neighboring Baltic Finnic languages, chiefly Livonian: interrogative particle *vai* (Liv. *või*), *māja* ‘house’ (Est. *maja*); and also by German: Middle Low German during the Hanseatic period and New High German thereafter (*brīvs* ‘free’: MLG *vrī*, *ün* ‘and’: MLG *un*, *brīlles* ‘eyeglasses’ NHG *die Brille*). Nevertheless, Latvian preserves a number of Indo-European archaisms not found elsewhere in Baltic: *asins* ‘blood,’ *agrs* ‘early.’

Bibliography

- Dini P (2000). *Baltų kalbos: lyginamoji istorija*. Vilnius: Mokslo ir enciklopedijų leidybos institutas.
- Endzelin J (1922). *Lettische Grammatik*. Riga: A. Gulbis.
- Fennell T G & Giesen H (1980). *A grammar of modern Latvian* (3 vols). The Hague: Mouton.
- Forssman B (2001). *Lettische grammatik*. Dettelbach: J. H. Röhl.
- Gäters A (1977). *Die lettische Sprache und Ihre Dialekte*. The Hague: Mouton.
- Kabelka J (1982). *Baltų filologijos įvadas*. Vilnius: Mokslas.
- Mathiassen T (1997). *A short grammar of Latvian*. Columbus, OH: Slavica.
- Milenbahs K & Endzelins J (1923–1932). *Latviešu valodas vārdnīca* (4 vols). Riga: Kultūras fonda izdevums.
- Nau N (1998). *Latvian*. Munich and Newcastle: Lincom Europa.
- Rūķe-Draviņa V (1977). *The standardization process in Latvian: 16th century to the present*. Stockholm: Almqvist & Wiksell International.

Lithuanian

S Young, University of Maryland Baltimore County, Baltimore, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Lithuanian (*Lietuvių kalbà*) is the native language of some 2.9 million speakers in the Republic of Lithuania. Together with Latvian, it forms East Baltic, the sole remaining branch of the Baltic family of Indo-European languages. There are two major dialects of Lithuanian, the more conservative and territorially greater Aukštaitic (*aukštaičių tarmė*) and the more innovating Žemaitic (*žemaičių tarmė*; Samogitian), spoken in the northwest quarter of Lithuania. The standard language is based on the speech of the southwest Aukštaitic region, bordering former East Prussia.

The Written Language

The Writing Tradition in East Prussia

Lithuanian is attested in written form from the 16th century, in three varieties of the Aukštaitic dialect; the earlier texts are chiefly translated and original religious literature. Book publication in Lithuanian began earliest in German East Prussia (which had a substantial Lithuanian population) in connection with the spread of the Reformation. The first work published in Lithuanian is a 1547 translation of a Lutheran catechism by Martynas Mažvydas (Martinus Masvidius). The foreword begins with a personal appeal to the reader, *Bralei seseris imkiet mani ir skaitikiet* ‘Brothers, sisters, take me and read me’. The language reflects Mažvydas’s native south Žemaitic dialect, with Aukštaitic elements. Subsequent Lithuanian publications in East Prussia are written in an increasingly normalized variety of the local west Aukštaitic dialect, codified in Daniel Klein’s 1653 *Grammatika Litvanica*, the first grammar of Lithuanian.

The Writing Tradition in the Grand Duchy

In the Catholic Grand Duchy of Lithuania, two writing traditions took root, one based on the East Aukštaitic dialect of the capital, Vilnius, and the other representing the Central Aukštaitic dialect of the Kedainiai area. The latter served as the medium for the earliest Lithuanian publications in the Grand Duchy, i.e., Mykalojus Daukša’s 1595 translation of Jacobus Ledisma’s popular Catholic catechism and his lengthy 1599 translation of Jakub Wujek’s collection of sermons, the *Postilla Catholicka*. Although these were translations, the language of these works

is relatively natural and had considerable influence on the later cultivation of Lithuanian. Daukša’s works are also the first accented texts in Lithuanian, and as such are of particular importance for the study of the historical prosody of the language.

The National Standard Language

The increasing polonization of the Grand Duchy’s nobility and educated classes led in the 18th century to a decline in the Central and East Aukštaitic writing traditions (the latter eventually disappeared). The present-day standard language has its roots in the late 19th century, and is based on the dialect of the southern West Aukštaitic region, in which the speakers are traditionally called *suvalkiečiai*. Several factors stand out in the establishment of this dialect as the national standard: the prior literary tradition of the virtually identical Aukštaitic dialect of neighboring East Prussia; the authority of the 19th-century Lithuanian grammars of A. Schleicher and F. Kurschat, which described the same Prussian Lithuanian speech; and the normative influence of late 19th- to early 20th-century newspapers such as *Aušra* (*The Dawn*) and *Varpas* (*The Bell*), which had many writers and editors (in particular Jonas Jablonskis) who came from the southwest Aukštaitic dialect area.

Phonology

Prosodic Features

Standard Lithuanian has free stress, which may alternate between a stem and ending within a grammatical paradigm, as in *dukrà* (nominative), *dùkrą* (accusative) ‘daughter’; *sakaũ* ‘I say’, and *sàko* ‘he says’. There are four such stress patterns for nouns and two for verbs. Stressed long vowels and diphthongs (including sequences of vowel plus tautosyllabic resonant) distinguish two phonemic contour tones, traditionally referred to as acute (´) and circumflex (˘), as in *šáu!* ‘shoot.IMP’ vs. *šaũ!* ‘shout.IMP’. Short stressed vowels are marked with a grave accent (`). The tones are conventionally indicated in dictionaries and linguistic works; otherwise, they are not represented.

According to the norms of the standard language, acute tone (*virtapradė priegaidė*) is realized with a falling tonal contour, whereas circumflex tone (*virtagalė priegaidė*) is level or rising. The tonal opposition is clearest on diphthongs; in the urban colloquial language, the distinction is becoming neutralized on long vowels. The phonetic realization of the two tones differs dialectally; in particular, the acute tone

of northwest Žemaitic speech incorporates a glottal stop (*laužtinė priegaidė* ‘broken tone’).

The Vowel System

Vowel length is distinctive in Lithuanian. The rather open short vowel phonemes /i [i], u [u], e [ɛ], a [a]/ (orthographically *i, u, e, a*) are inherited from proto-Baltic and also result from an early Lithuanian shortening of final long vowels under acute tone (compare *tà* ‘this. NOM SG FEM’ with Latvian *tā*, having the Latvian reflex of acute). In addition, a short /ɔ/ (spelled *o*) is found in words of foreign origin.

The long vowel phonemes /i:, u:, e:, o:, æ:, a:/ (orthographically *y/i, ū/u, è, o, e, q*) also have two sources. Inherited length is represented by the spellings *y, ū, è, o* [*< *ā*], as in *gývas* (**gī-*) ‘alive’, *búti* (**bū-*) ‘to be’, *sėti* (**sē-*) ‘to sow’, and *brólis* (Latvian *brālis*) ‘brother’, whereas *i, u, e, and q* develop from sequences of vowel plus tautosyllabic *n*, when not before a stop (where they are preserved). Original *V + n* sequences were first replaced by long nasalized vowels, marked in the earlier texts by a hook under the corresponding vowel graph. These vowels were eventually denasalized, although the orthography still reflects the earlier practice, as in *ĩ* [i:] (< **in*) ‘to, into’, *siūsti* [s’iūst’i] (< **siuñt-*) ‘to send’, *tęsti* [t’iæ:st’i] (< **teñs-*) ‘to continue’, and *žąsis* [ž’a:s’is] (< **žañs-*) ‘goose’. Both long and short *a* are fronted to /æ:/ and /ɛ/, respectively, after a palatalized consonant or *j*, as in *gilią* ‘deep.ACC SG FEM’ = *gile* ‘acorn.ACC SG FEM’, both [g’i’l’æ:]; and *giliąs* ‘deep.ACC PL FEM’ = *gilės* ‘acorns.ACC PL FEM’, both [g’i’l’ɛs].

Short *e* and *a* are automatically lengthened under stress in most nonfinal syllables to [æ:] and [a:], with concomitant circumflex tone. This phonetic vowel length is not indicated orthographically, i.e., *ledas* [l’æ:das] ‘ice’ and *vakaras* [v’ā:karas] ‘evening’ (compare Latvian *ledus* and *vakars*, having short *e* and *a*). Also included in the inventory of Lithuanian vowels are the diphthongs *ie* and *uo*, which arose from East Baltic **ē* (< **ei*) and **ō* (< **o*). These diphthongs, which function as long vowels, begin with a high vowel and end with a lower, more central vowel (ɨ, ʊ), as in *dienà* (**dein-*) ‘day’ and *dúona* (**dōn-*) ‘bread’, phonetically [diə’na] and [dʊ’na].

The Consonant System

Lithuanian alone among the Baltic languages preserves distinct reflexes (*j* and *ʒ*) of the Indo-European palatovelars; in Latvian and Old Prussian as well as in Slavic, these have merged with *s* and *z*; as in *šuō* ‘dog’ and *žemė* ‘earth’ (Latvian *suns* and *zeme*).

A characteristic feature of the Lithuanian consonant system is the phonemic opposition of palatalized and nonpalatalized consonants before back vowels (palatalization is automatic before front vowels). These palatalized consonants are the product of Baltic consonant + *j* sequences, which are still preserved in the case of labial stops in word-initial position, as in *pjáuti* [p’jæut’i] ‘to cut’ and *bjaurūs* [b’jæu’rus] ‘ugly’. Earlier sequences of dental stop + *j* have developed into the affricates [tʃʲ] and [dʒʲ], orthographically *č(i)* and *ž(i)*, as in *čià* (**tja*) ‘here’ and *mėdžias* (**medjas*) ‘woods’. The distinctive palatalization of the remaining consonants is marked orthographically by a following *i*, as in *siūti* [s’iūti] ‘to sew’ and *liáudis* [l’i’áud’is] ‘people’.

Morphosyntactic Features

The Noun

Lithuanian has preserved the Baltic stem classes and their declensional endings rather well, giving the noun a relatively archaic appearance. Six case forms are distinguished in the singular and plural (a dual is attested for certain case forms in dialects and older texts): (1) nominative (for example, in the singular, the *o*-stem *nāmas* ‘house’, *ā*-stem *rankà* ‘hand’, *i*-stem *akis* ‘eye’, and *u*-stem *sūnūs* ‘son’), (2) genitive (*nāmo, rankos, akiės, and sūnaūs*), (3) dative (*nāmui, rankai, ākiai, and sūnui*), (4) accusative (*nāma, ranką, ākį, and sūnų*), (5) instrumental (*namū, rankà, akimi, and sūnumi*), and (6) locative (*namè, rankojè, akyjè, and sūnujè*). In addition, a special vocative form is used in the singular, as in *Jōnai!* (nominative *Jōnas* ‘John’) and *Birùte!* (nominative *Birùtė*).

The adnominal genitive is typically preposed, as in *tėvo nāmas* ‘of-father house’ (‘father’s house’) and *lietūvių kalbà* ‘of-Lithuanians language’ (‘Lithuanian language’). The genitive also occurs in partitive expressions, both positive, as in *miškè yrà vilkū* ‘in-the-woods there-are wolves (genitive)’, and negative, as in *nèrà žuviès* ‘there-is-no fish (genitive)’. The locative case is used without a preposition, as in *Vilniuje* ‘in Vilnius’; historically this represents an inessive, the remnant of a more complex system of local cases formed with postpositions. These cases included an adessive, illative, and allative, some of which (particularly the illative) are still found dialectally.

Nouns are marked for gender (masculine and feminine) through distinctive desinences and adjectival concord, as in *gėras tėvas* ‘good father’ and *gerà mótina* ‘good mother’. In an innovation shared with

Latvian, proto-Baltic neuter gender was lost; neuters typically became masculine, as in *šīēnas* ‘hay’ (masc.) and Old Church Slavonic *sěno* (neut.). The category of definiteness is marked within adjectives by the historical affixation of a pronominal *-ī-* element to the indefinite form, as in indefinite *naūjas* (masc.), *naujā* (fem.) ‘new’ vs. definite *naujāsis* (i.e., *naujas* + *jis*) (masc.), *naujōji* (fem.).

The Verb

The Lithuanian verb marks present, past, and future tense forms. The present tense has three conjugation patterns, illustrated by *lipti* ‘to climb’ (first conjugation, stem in *a*), *mylēti* ‘to love’ (second conjugation, stem in *i*), and *skaityti* ‘to read’ (third conjugation, stem in *o*): 1SG (*aš*) *lipū*, *myliu*, *skaitau*; 1PL (*mes*) *lipame*, *mylime*, *skaitome*; 2SG (*tu*) *lipi*, *myli*, *skaitai*; 2PL (*jūs*) *lipate*, *mylite*, *skaitote*; 3SG/PL (*jis/jie*) *lipa*, *myli*, *skaito*. The past tense has two patterns, illustrated by *lipti* ‘to climb’ (stem in *o*) and *skaityti* ‘to read’ (stem in *ē*): 1SG (*aš*) *lipau* *skaičiau*; 1PL (*mes*) *lipome*, *skaitėme*; 2SG (*tu*) *lipai*, *skaitėi*; 2PL (*jūs*) *lipote*, *skaitėte*; and 3SG/PL (*jis/jie*) *lipo*, *skaitė*. A frequentative past is formed by adding the suffix *-dav-* (plus *o*-stem endings) to the infinitive stem, as in *jis skaitydavo* ‘he used to read, would read’ (*skaityti* ‘to read’). The future is formed by adding *-s-* and the present-tense person endings to the infinitive stem, *lipti* ‘to climb’: 1SG (*aš*) *lipsiu*, 1PL (*mes*) *lipsime*, 2SG (*tu*) *lipsi*, 2PL (*jūs*) *lipsite*, and 3SG/PL (*jis, jie*) *lips*. As the various examples demonstrate, number is not marked in the third person, a characteristic feature of Baltic.

Lithuanian shows a fondness for participles and gerunds, in both colloquial and written styles. Among the more typical participles (which decline like adjectives) are the present active (*rašys*, stem *rāšant-*, infinitive *rašyti* ‘to write’), past active ([*pa*]*rāšęs*, stem ([*pa*]*rāšius-*), present passive (*rāšomas*), and past passive (*parašytas*). The past active participle is used, together with a finite form of the verb ‘to be,’ to form a system of perfect tenses, as in *aš esu* (*pa*)*rāšęs* (masc.)/(*pa*)*rāšiusi* (fem.) ‘I have written’ and *aš buvau* (*pa*)*rāšęs* (masc.)/(*pa*)*rāšiusi* (fem.) ‘I had written’. Passive constructions are formed with the verb ‘to be’ and the corresponding passive

participle, as in *knygà būvo rāšoma* ‘the book was being written’ and *knygà būvo parašyta* ‘the book was/had been written’. The language retains a reflex of an earlier dative absolute in gerundive constructions such as *sáulei tēkant* ‘as the sun is/was rising’ (‘to-the-sun rising’).

Lexicon

Lithuanian has long felt the lexical influence of neighboring Slavic languages. Early East Slavic borrowings into Lithuanian include *tuřgus* ‘market’ and *krikřtas* ‘baptism’ (Old Russian *tōrgō* ‘market’ and *kr̥stō* ‘cross’). Among East Slavic borrowings from the time of the Grand Duchy are *knygà* ‘book’ and *blýnas* ‘pancake’ (Russian *kníga* and *blin*). A significant number of Polish borrowings began to appear in Lithuanian in the 17th and 18th centuries, among them *arbatà* ‘tea’ and *cūkrus* ‘sugar’ (Polish *herbata* and *cukier*). Since the late 19th century, language reformers have succeeded in replacing many earlier borrowings with native words known from dialects and Old Lithuanian texts; for example, the native *laikas* ‘time’ was normalized in place of *čėsas* (Russian *čas*), and *pasáulė* ‘world’, in place of *svietas* (Russian *svet*). During this time, a number of neologisms took root, such as *akiniaĩ* ‘eyeglasses’ (*akis* ‘eye’) and *mokyklà* ‘school’ (*mok-* ‘teach, learn’ + *-ykl-* ‘place where’).

Bibliography

- Ambrasas V (ed.) (1997). *Lithuanian grammar*. Vilnius: Baltos Lankos.
- Dini P (2000). *Baltų kalbos: lyginamoji istorija*. Vilnius: Mokslo ir Enciklopedijų Leidybos Institutas.
- Kabelka J (1982). *Baltų filologijos įvadas*. Vilnius: Mokslas.
- Mathiassen T (1996). *A short grammar of Lithuanian*. Columbus, OH: Slavica.
- Palionis J (1979). *Lietuvių literatūrinės kalbos istorija*. Vilnius: Mokslas.
- Senn A (1966). *Handbuch der litauischen Sprache (vol. 1): Grammatik*. Heidelberg: C. Winter.
- Zinkevičius Z (1966). *Lietuvių dialektologija*. Vilnius: Mintis.
- Zinkevičius Z (1984–1995). *Lietuvių kalbos istorija* (6 vols.). Vilnius: Mokslas.

Long-Range Comparison: Methodological Disputes

L Campbell, University of Utah, Salt Lake City, UT, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

How are languages shown to be related to one another? Proposals of distant linguistic kinship such as Amerind, Nostratic, Eurasiatic, and Proto-World have received much attention in recent years, although these same proposals are rejected by a majority of practicing historical linguists. This has resulted in vigorous disputes about the methods for investigating remote relationships among languages as yet not known to be related. Some enthusiasts of long-range relationships, disappointed that proposed language connections they favor have not been accepted, have at times responded bitterly, for example charging that these rejections are just “clumsy and dishonest attempts to discredit deep reconstructions” (Shevoroshkin, 1989: 7), and that “very few [critics of long-range proposals] have ever bothered to examine the evidence first-hand... To really screw up classification you almost have to have a Ph.D. in historical linguistics” (Ruhlen, 1994: viii). The strong rhetoric is not all one-sided:

At a different level – which transcends scientific worth to such an extent that it is at the fringe of idiocy – there have in recent years been promulgated a number of far-fetched ideas concerning ‘long-distance relationships’, such as ‘Nostratic’, ‘Sino-Caucasian’, and ‘Amerind’. (Dixon, 2002: 23)

This article explains these disputes.

Hypothesized Long-Range Relationships

The list in Table 1 of the better-known hypotheses that would group together languages that are not yet known to be related gives an idea of what is at issue. (None of the proposed genetic relationships in this list has been demonstrated yet, even though some are repeated frequently.)

Methods

Scholars agree that a successful demonstration of linguistic kinship depends on adequate methods, but disagree about what these methods are. Hence discussions of methodology assume a central role in considerations of long-range comparisons. Therefore, the methodological principles and criteria considered important for investigating proposals of distant genetic relationship are surveyed here.

In practice, the successful methods for establishing distant linguistic affinity have not been different from those used to establish any family relationship, close or distant. The comparative method has always been the primary tool. Because the methods for distant relationships are not different from those for more closely related languages, we encounter a continuum from established families (e.g., Indo-European, Finno-Ugric, Mayan, Bantu), to more distant but solidly demonstrated relationships (e.g., Uralic, Siouan-Catawban, Benue-Congo), to plausible but inconclusive hypotheses (e.g., Indo-Uralic, Proto-Australian, Macro-Mayan, Niger-Congo), to doubtful but not implausible ones (e.g., Altaic, Austro-Tai, Eskimo-Uralic, Nilo-Saharan), and on to virtually impossible proposals (e.g., Basque-Sino-Tibetan-NaDene, Indo-Pacific Mayan-Turkic, Miwok-Uralic, Niger-Saharan).

Table 1 Proposals of distant genetic relationships among languages

| |
|--|
| Altaic (Turkic, Tungusic, Mongolian, sometimes Japanese, Korean) |
| Amerind (uniting all Native American language families, except Eskimo-Aleut and Na-Dene) |
| Dene-Sino Tibetan (Athabaskan [or Na-Dene] and Sino-Tibetan) |
| Austriac (Austro-Asiatic with Austronesian) |
| Austro-Tai (Japanese-Austro-Thai) |
| Basque-Caucasian, Basque-Sino Tibetan-NaDene |
| Dravidian-Uralic |
| Eskimo and Indo-European |
| Eskimo-Uralic |
| Eurasiatic (Indo-European, Uralic, Eskimo-Aleut, Ainu, several others) |
| Hokan (grouping numerous American Indian families and isolates) |
| Indo-European and Afroasiatic, Indo-European and Semitic |
| Indo-Pacific (grouping the non-Austronesian languages of the Pacific: all Papuan families, Tasmanian, languages of the Andaman Islands) |
| Japanese-Austronesian |
| Khoisan (grouping most non-Bantu African click languages, an areal grouping, not a genetic one) |
| Macro-Siouan (Siouan, Iroquoian, Caddoan, sometimes Yuchi) |
| Maya-Chipayan (Mayan, Uru-Chipayan of Bolivia) |
| Na-Dene (Eyak-Athabaskan, Tlinglit, Haida; Haida is highly disputed) |
| Niger-Kordofanian (Niger-Congo) (grouping Mande, Kru, Kwa, Benue-Congo [of which Bantu is a branch], Gur, Adamawa-Ubangi, Kordofanian, etc.) |
| Nilo-Saharan (most of the African languages not otherwise classified with one of Greenberg's other three African macrofamilies) |
| Nostratic (Indo-European, Uralic, Altaic, Kartvelian, Dravidian, Afroasiatic; some add others) |
| Penutian (grouping numerous American Indian families and isolates) |
| Proto-Australian (all Australian families) |
| Proto-World (uniting all the world's languages) |
| Ural-Altaic (Uralic and Altaic) |

It is difficult on the basis of standard methods to segment this continuum so that plausible proposals based on legitimate procedures fall sharply on one side, distinguished from obviously unlikely hypotheses on the other side. This leads to disagreements, even by those who profess allegiance to the same methods.

A firm understanding of methodology becomes crucial if supporters of fringe proposals can pretend to apply the same methods as those employed for more plausible ones. For this reason, careful evaluation of the evidence presented on behalf of any proposed distant linguistic relationship and of the methods employed is called for.

Throughout history, the criteria employed in both pronouncements about method and in actual practice for establishing language families consistently included evidence from three sources: basic vocabulary, grammatical evidence (especially morphological), and sound correspondences. Hoenigswald (1990: 119–120) summarized the points upon which 17th- and 18th-century linguistic scholars agreed:

There was ... “a concept of the development of languages into dialects and of dialects into new independent languages” ... and ... “an insistence that not a few random items, but a large number of words from the **basic vocabulary** should form the basis of comparison” ... the doctrine that ‘**grammar**’ is even more important than words; ... the idea that for an etymology to be valid the **differences in sound** – or in ‘letters’ – **must recur** [emphasis added, LC].

These criteria figured prominently in nearly all demonstrations of language families in the past, making them important also to today’s practice. The methods and criteria generally thought necessary for reliable long-range comparison are surveyed in what follows. (See Campbell, 2003; Campbell, 1997a: 206–259 for details.)

Lexical Comparison

Throughout history, word comparisons have been employed as evidence of language family relationship, but, given a small collection of likely-looking words, how can we determine whether they are really the residue of common origin and not due to chance or some other factor? Lexical comparisons by themselves are seldom convincing without additional support from other criteria.

Basic Vocabulary Most scholars require that basic vocabulary be part of the supporting evidence for any distant genetic relationship. Basic vocabulary is generally understood to include terms for body parts, close kinship, frequently encountered aspects of the natural world (mountain, river, cloud), and low numbers. Basic vocabulary is generally resistant to

borrowing, so comparisons involving basic vocabulary items are less likely to be due to diffusion and stand a better chance of being inherited from a common ancestor than other kinds of vocabulary. Still, basic vocabulary can also be borrowed – though infrequently – so that its role as a safeguard against borrowing is not foolproof.

Glottochronology Glottochronology, now mostly abandoned, aimed at assigning dates to the split up of related languages; it has been employed in long-range comparisons. It depends on basic, relatively culture-free vocabulary, but all its basic assumptions have been challenged (including the existence of culture-free vocabulary). Most tellingly, it does not find or test distant genetic relationships, but rather it **assumes** that the languages compared are related and proceeds to attach a date based on the number of core-vocabulary words that are considered similar among the languages compared. This, then, is no method for determining whether languages are related.

Multilateral Comparison The best-known approach that relies on inspectional resemblances among words is Joseph Greenberg’s ‘multilateral (or mass) comparison.’ It is based on “looking at ... many languages across a few words” rather than “at a few languages across many words” (Greenberg, 1987: 23). The lexical similarities determined by superficial visual inspection that are shared ‘across many languages’ alone are taken as sufficient evidence for genetic relationship. This approach stops where others begin, at assembling lexical similarities. These inspectional resemblances must be investigated to determine why they are similar, whether the similarity is due to inheritance from a common ancestor (genetic relationship), or to borrowing, accident, onomatopoeia, sound symbolism, or nursery formations – nongenetic factors. Since multilateral comparison does not do this, its results are controversial and rejected by most mainstream historical linguists.

No technique that relies on inspectional similarities in vocabulary alone has proven adequate for establishing family relationships.

Sound Correspondences

Nearly all scholars consider regular sound correspondences strong evidence of genetic affinity. Correspondences do not necessarily involve similar sounds. The sounds that are equated in proposals of remote relationship are typically similar, often identical, although such identities are not so frequent among the daughter languages of well-established language families. The sound changes that lead to such

nonidentical correspondences often make cognate words not apparent. These true but nonobvious cognates are missed by methods that seek only superficial resemblance, for example: French *cinq*/Russian *pjatj*/Armenian *hing*/English *five* (all derived by straightforward changes from original Indo-European **penkwe*- ‘five’); French *boeuf*/English *cow* (both from Proto-Indo-European **gwou*- ‘cow’). The words in these cognate sets are not visually similar, but they exhibit regular sound correspondences among the cognates.

Though extremely important and valuable, the criterion of sound correspondences can be misapplied. Sometimes regularly corresponding sounds are found in loans. By Grimm’s law, real French-English cognates should exhibit the correspondence *p* : *f*, as in the cognates *père*/father, *pie*/foot, *pour*/for. However, French and English appear to correspond *p* : *p* in cases where English has borrowed from French or Latin, as in *paternel*/paternal, *pié*/pedestal/pedestal, *per*/per. Since English has many such loans, examples of the bogus *p* : *p* sound correspondence are not hard to find. In comparing languages not yet known to be related, we must be cautious of the problem of seeming correspondences in undetected loans. Sound correspondences in basic vocabulary help, since basic vocabulary is borrowed only infrequently.

Some nongenuine sound correspondences can come from accidentally similar words. Languages share some vocabulary by sheer accident, for example: Proto-Je **niw* ‘new’/English *new*; Kaqchikel *mes* ‘mess’/English *mess*; Maori *kuri* ‘dog’/English *cur*; Lake Miwok *hóllu* ‘hollow’/English *hollow*; Gbaya *be* ‘to be’/English *be*. Other unreal sound correspondences can come from wide semantic latitude in proposed cognates, when phonetically similar but semantically disparate forms are equated. For example, if we compare Pipil (Uto-Aztecan) *teki* ‘to cut’/Finnish (Uralic) *teki* ‘made’, *tukat* ‘spider’/tukat ‘hairs’, etc., we note a recurrence of a *t* : *t* and a *k* : *k* correspondence. However, the phonetic correspondences in these word pairs are accidental – it is always possible to find phonetically similar words among languages if their meanings are ignored. With too much semantic leeway among compared forms, spurious correspondences such as the Pipil-Finnish *t* : *t* and *k* : *k* turn up. Unfortunately, wide semantic latitude is very common in cases of long-range comparison. Additional noninherited phonetic similarities crop up when onomatopoeic, sound-symbolic, and nursery forms are compared. A set of proposed cognates involving a combination of loans, chance enhanced by semantic latitude, onomatopoeia, and such factors can exhibit false sound correspondences.

For this reason, some proposed remote relationships that purportedly are based on regular sound correspondences nevertheless fail to be convincing.

Grammatical Evidence

Scholars throughout linguistic history have considered morphological evidence important for establishing language families. Many favor ‘shared aberrancy’ (‘submerged features,’ ‘morphological peculiarities,’ ‘arbitrary associations’). For example, the Algonquian-Ritwan hypothesis, which groups Wiyot and Yurok (both of California) with the Algonquian family, was controversial, but morphological evidence such as the following comparison of Proto-Central-Algonquian (PCA) and Wiyot helped to confirm the relationship:

Proto-Central-Algonquian **ne* + **ehkw*-
= **netehkw*-‘my louse’
Wiyot *du*- + *híkw* = *duhíkw* ‘my louse’ (Teeter, 1964: 1029)

Proto-Central-Algonquian inserts *-t* between a possessive pronominal prefix and a vowel-initial root, while Wiyot inserts *-t-* between possessive prefixes and a root beginning in *hV* (with the loss of the *h* in this process). There is no phonetic reason why *t* should be added in this environment; this is so unusual it is not likely to be shared by borrowing or accident. Inheritance from a common ancestor that had this peculiarity is more likely, and this is confirmed by other evidence in these languages. Another often repeated example of shared aberrancy is the suppletive agreement between English *good/better/best* and German *gut/besser/best*, where examples such as this are held to have probative value for showing languages are related.

Morphological correspondences of the ‘shared aberrancy’ type are an important source of evidence for distant genetic relationships.

Borrowing

Diffusion is a source of nongenetic similarity among languages that can complicate evidence for remote relationships. For example, the controversial ‘Chibchan-Paezan’ hypothesis (grouping several South American language families, part of ‘Amerind’) has the proposed cognate set ‘axe’ with words from only four of the many languages involved, but two of these are loans: Cuitlatec *navaxo* ‘knife’, from Spanish *navajo* ‘knife, razor’, and Tunebo *baxita* ‘machete’, from Spanish *machete* (Tunebo has nasal consonants only before nasal vowels, hence *b* substitutes for Spanish *m*) (Greenberg, 1987: 108). When two of the four pieces of evidence are borrowings, the putative ‘axe’ cognate must be abandoned. Examples

such as this are not uncommon in proposals of distant genetic relationship.

Semantic Constraints

It is dangerous to present phonetically similar forms with different meanings as potential evidence of remote genetic relationship, assuming semantic shifts have taken place. Of course meaning can shift, but in hypotheses of remote relationship the assumed semantic shifts cannot be documented, and the greater the semantic latitude permitted in compared forms, the easier it is to find phonetically similar forms that have no historical connection (as in the Pipil-Finnish examples, above). When semantically nonequivalent forms are compared, chance phonetic similarity is greatly increased. Within families where the languages are known to be related, etymologies are not accepted unless an explicit account of any assumed semantic changes can be provided. The problem of excessive semantic permissiveness is one of the most common and most serious in long-range comparisons, for example, sets cited for Nostratic compare forms meaning 'lip/mushroom/soft outgrowth', 'grow up/become/tree/be', for Amerind hypothesis 'excrement/night/grass', 'child/copulate/son/girl/boy/tender/bear/small'. It is for reasons such as this that these proposals of more remote linguistic relationship are disputed.

Onomatopoeia

Onomatopoeic words imitate the real-world sound associated with their meanings. They may be similar in different languages because they have independently approximated the sounds of nature, not because they share any common history. A way to reduce the sound-imitative problem is to omit from long-range comparisons any word which cross-linguistically frequently has similar imitative form, for example 'blow', 'breathe', 'suck', 'laugh', 'cough', 'sneeze', 'break/cut/chop/split', 'cricket', 'crow' (bird names in general), 'frog/toad', 'lungs', 'baby/infant', 'beat/hit/pound', 'call/shout', 'choke', 'cry', 'drip/drop', 'hiccough', 'kiss', 'shoot', 'snore', 'spit', and 'whistle'. Unfortunately, examples of onomatopoeic words are frequent in proposals of distant genetic relationships.

Nursery Forms

Nursery words (the 'mama-nana-papa-dada-caca' sort) should be avoided, since they typically share a high degree of cross-linguistic similarity that is not due to common ancestry. Nevertheless, examples of nursery words are frequent in cases of long range comparison. The words involved are typically 'mother', 'father', 'grandmother', 'grandfather', and often 'brother', 'sister', 'aunt', and 'uncle', and have shapes

like *mama, nana, papa, baba, tata, dada*. Jakobson (1962[1960]: 542–543) explained the cross-linguistic nongenetic similarity among nursery forms. Nursery words provide no reliable support for genetic relationship.

Short Forms and Unmatched Segments

How long proposed cognates are and the number of matched sounds within them are important, since the greater the number of matching sounds in a proposed cognate set, the less likely it is that accident accounts for the similarity. Monosyllabic words (CV, VC, V) are so short that their similarity to forms in other languages could also easily be due to chance. If only one or two sounds of longer forms are matched, chance may explain the similarity. Such comparisons are not persuasive.

Chance Similarities

Chance (accident) is another possible explanation for similarities in compared languages and needs to be avoided. The potential for accidental matching increases dramatically when one leaves the realm of basic vocabulary, when one increases the pool of words from which potential cognates are sought, and when one permits the semantics of compared forms to vary even slightly (Ringe, 1992: 5).

Cases of similar but noncognate words are well-known, for example French *feu* and German *Feuer* 'fire', English *much* and Spanish *mucho* 'much'. The phonetic similarity in these basic vocabulary items is due to accidental convergence due to the sound changes that they have undergone, not to inheritance from any common word in the proto language. That originally distinct forms in different languages can become similar due to sound changes is not surprising, since even within a single language originally distinct words can converge due to sound changes, for example, English *liellie* (from Proto-Germanic **ligjan* 'to lie, lay')/**leugan* 'to tell a lie').

Sound-Meaning Isomorphism

Only comparisons which involve both sound and meaning together are permitted. Similarities in sound alone (for example, the presence of tones in compared languages) or in meaning alone (for example, grammatical gender in languages compared) are not reliable, since they can develop independently of genetic relationship, due to diffusion, accident, and typological tendencies.

Only Linguistic Evidence

Only linguistic information, no nonlinguistic consideration, is permitted as evidence of distant genetic

relationship. Shared cultural traits, mythology, folklore, technologies, and gene pools must be eliminated from arguments for linguistic relationship. The wisdom of this is seen in face of the many strange proposals based on nonlinguistic evidence. For example, some earlier African classifications proposed that Ari (Omotic) belongs to either Nilo-Saharan or Sudanic 'because the Ari people are Negroes' ('racial' evidence), that Moru and Madi belong to Sudanic because they are located in central Africa (geographical evidence), or that Fula is Hamitic because its speakers herd cattle, are Moslems (cultural evidence), and are tall and Caucasoid (physical attributes) (Fleming, 1987: 207). Clearly the language one speaks does not deterministically depend on one's cultural and biological connections.

Erroneous Morphological Analysis

Where compared words are analyzed with more than one morpheme, it is necessary to show that the segmented morphemes in fact exist in language. Unfortunately, unmotivated morphological divisions are frequent in proposals of remote relationship. Often, a morpheme boundary is inserted where none is justified, as for example, the arbitrarily segmented Tunebo 'machete' as *baxi-ta* (borrowed from Spanish *machete*, and contains no morpheme boundary). This false morphological segmentation falsely makes the Tunebo word appear more similar to the other proposed cognates, Cabecar *bak* and Andaqui *boxo-(ka)* 'axe' (Greenberg, 1987: 108).

Undetected morpheme divisions are also a problem. An example from the Amerind hypothesis compares Tzotzil *ti?il* 'hole' with Lake Miwok *talokh* 'hole', Atakapa *tol* 'anus', Totonac *tan* 'buttocks', Takelma *telkan* 'buttocks' (Greenberg, 1987: 152); however, the Tzotzil form is *ti?-il*, from *ti?* 'mouth' + *-il* 'indefinite possessive suffix', meaning 'edge, border, lips, mouth', but not 'hole'. The appropriate comparison *ti?* 'mouth' bears no particular resemblance to the other forms with which it is compared.

Spurious Forms

Another problem is that of nonexistent or erroneous 'data' from 'bookkeeping' problems and 'scribal' errors. For example, for the Mayan-MixeZoquean hypothesis (Brown and Witkowski, 1979), Mixe-Zoquean words meaning 'shell' were compared with K'iche' (Mayan) *sak*, said to mean 'lobster', actually 'grasshopper' – a misunderstanding of Spanish *langosta*, which in Guatemala (where K'iche' is spoken) means 'grasshopper', but 'lobster' in other varieties of Spanish. A comparison of 'shell' and 'grasshopper' makes no sense. Errors of this sort can be serious;

for example, in the Amerind hypothesis (Greenberg, 1987) none of the words given as Quapaw are in fact from Quapaw; all are from Biloxi and Ofo; none of the words given as Proto-Mayan are from Proto-Mayan, rather from Proto-K'ichean.

Given the disputes about proposed distant genetic relationships, these methodological principles for long-range comparison are extremely important. Research on possible distant genetic relationships that does not conform to these methodological principles and cautions will remain inconclusive.

Some Examples of Long-Range Proposals

It will be instructive to look briefly at some specific proposals to see why most mainstream historical linguists do not accept these hypotheses. Space does not permit full evaluation, but references are given for more detail.

Altaic

The Altaic hypothesis would group Turkic, Mongolian, and Tungusic; some versions also include Korean and Japanese. While 'Altaic' is repeated in encyclopedias, most leading 'Altaicists' have abandoned the hypothesis. The most serious problems are the extensive borrowing among the 'Altaic' languages, lack of convincing cognates, lack of basic vocabulary, extensive areal diffusion, problems with the putative sound correspondences, and reliance on typologically commonplace traits. The shared 'Altaic' traits include vowel harmony, relatively simple phoneme inventories, agglutination, suffixing, (S)OV word order (and postpositions), no verb 'to have' for possession, no articles or gender, and nonmain clauses in nonfinite (participial) constructions. However, these shared features are commonplace typological traits, and thus are not good evidence of genetic relationship because they can easily develop independently in unrelated languages. These 'Altaic' features are also areal traits, shared by a number of languages in surrounding regions, thus perhaps due to diffusion. Similarities in the first and second person pronoun paradigms have impressed proponents of Altaic, although critics point out that pronouns are borrowed far more frequently than proponents acknowledge and pronoun patterns of the type cited for Altaic are also not unusual nor unexpected cross-linguistically. In short, the evidence for genetic relationship has not been persuasive, explaining why so many reject the 'Altaic' hypothesis. (Campbell and Poser, 2008).

Nostratic

The Nostratic hypothesis as advanced in the 1960s by Illich-Svitych would group Indo-European, Uralic,

Altaic, Kartvelian, Dravidian, and Hamito-Semitic [later Afroasiatic], though other versions of the hypothesis would include various other languages. The sheer number of languages and many proposed cognates involved might make it seem difficult to evaluate Nostratic. Nevertheless, assessment is possible. With respect to the many putative cognate sets, assessment can concentrate on those cases considered the strongest by proponents of Nostratic (those of Dolgopolsky, 1986 and Kaiser and Shevoroshkin, 1988). Campbell (1998) shows that these strongest cases do not hold up and that the weaker sets are not persuasive (see below). We can easily determine to what extent the proposed reconstructions correspond to typological expectations, whether the proposed cognates are permissive in semantic associations, and when onomatopoeia, forms too short to deny chance, nursery forms, and the like are involved.

Illich-Svitych's version of Nostratic exhibits the following methodological problems. (See Campbell, 1998, 1999 for details.)

1. Descriptive forms. Illich-Svitych is forthright in labeling 26 of his 378 forms as 'descriptive,' meaning onomatopoeic, affective, or sound-symbolic, i.e., 7% of the total. There are 16 additional onomatopoeic, affective, or sound-symbolic forms, not so labeled, or a total of approximately 11%.
 2. Questionable cognates. Illich-Svitych himself indicates that 57 of the 378 sets are questionable (15%), signaled with a question mark. However, this number should be greatly increased, since in numerous forms Illich-Svitych signals problems in other ways, with slanted lines (/ /) for things not conforming to expectation, with question marks, and with upper-case letters in reconstructions to indicate uncertainties or ambiguities.
 3. Sets with only two families represented. One of Illich-Svitych's criteria was that only cognate sets with representatives from at least three of the six 'Nostratic' families would be considered as supportive. Nevertheless, 134 of the 378 sets involve forms from only two families (35%), questionable by Illich-Svitych's own criteria.
 4. Noncorresponding sound correspondences. Frequently, the forms presented as evidence of Nostratic do not exhibit the proposed sound correspondences, i.e., they have sounds at odds with those that would be required according to the Nostratic correspondence sets. Campbell (1998), looking mostly only at stops and only at the Indo-European and Uralic data, found 25 sets that did not follow the proposed Nostratic correspondences. There is another way in which
- Illich-Svitych's putative sound correspondences are not consistent with the standard comparative method. Several of the putative Nostratic sounds are not reflected by regular sound correspondences in the languages. For example, "in Kartv[e-lian] and Indo-European, the reflexes of Nostratic [**p*] are found to be unstable" (Illich-Svitych, 1990: 168). Nostratic forms beginning in **p* reveal that both the Indo-European and the Kartvelian forms arbitrarily begin with either **p* or **b*, but this is not regular sound change and is not sanctioned by the comparative method. Similarly, glottalization in Afroasiatic is said to occur "sporadically under other conditions still not clear" (Illich-Svitych, 1990: 168). In the correspondence sets, several of the languages are listed with multiple reflexes of a single Nostratic sound, but with no explanation of conditions under which the distinct reflexes might appear.
5. Short forms. Of Illich-Svitych's 378 forms, 57 (15%) involve short forms (CV, VC, C, or V), incapable of denying chance as an alternative explanation.
 6. Semantically nonequivalent forms. Some 55 cases (16%) involve comparisons of forms in the different languages that are fairly distinct semantically.
 7. Diffused forms. Given the history of central Eurasia, with much language contact, it is not at all surprising that some forms turn out to be borrowed. Several of the Nostratic cognates have words which have been identified by others as loans, including: 'sister-in-law', 'water', 'do', 'give', 'carry', 'lead', 'to do'/'put', 'husband's sister', to which we can add the following as probable or possible loans: 'conifer, branch, point', 'thorn'; 'poplar'; 'practice witchcraft'; 'deer'; 'vessel'; 'birch'; 'bird cherry'; 'honey', 'mead'; 'poplar'.
 8. Typological problems. Nostratic as traditionally reconstructed is typologically flawed. Counter to expectations, few Nostratic roots contain two voiceless stops; glottalized stops are considerably more frequent than their plain counterparts; Nostratic affricates change to a cluster of fricative + stop in Indo-European.
 9. Evaluation of the strongest lexical sets. An examination of the Nostratic sets held by proponents to be the strongest reveals serious problems with most. These include Dolgopolsky's (1986) 15 most stable lexemes. Most are questionable in one way or another according to the standard criteria for assessing proposals of remote linguistic kinship. In the Nostratic sets representing Dolgopolsky's 15 most stable glosses, four have problems with phonological correspondences; five

involve excessive semantic difference among the putative cognates; four have representatives in only two of the putative Nostratic families; two involve problems of morphological analysis; Illich-Svitych himself listed one as doubtful; and finally, one reflects the tendency to rely too heavily on Finnish when not supported by the historical evidence. All but two are challenged, and for these two the relevant forms needed for evaluation are not present. (See Campbell, 1998 for details.) These 'strong' cases are certainly not sufficiently robust to encourage faith in the proposed genetic relationship.

Once again, it is for reasons of this sort that most historical linguists reject Nostratic.

Amerind

Greenberg (1987) proposed that all Native American languages, except Na-Dene and Eskimo-Aleut languages, belong to single macro-family, Amerind, based on multilateral comparison (see above). Amerind is rejected by virtually all specialists in Native American languages and by the vast majority of historical linguists. Specialists maintain that valid methods do not at present permit reduction of Native American languages to fewer than about 150 independent language families and isolates. Amerind has been highly criticized on various grounds. There are exceedingly many errors in Greenberg's data: "the number of erroneous forms probably exceeds that of the correct forms" (Adelaar, 1989: 253). Where Greenberg stops – after assembling superficial similarities and declaring them due to common ancestry – is where other linguists begin. Since such similarities can be due to chance, borrowing, onomatopoeia, sound symbolism, nursery words (the *mama*, *papa*, *nana*, *dada*, *caca* sort), misanalysis, and much more, for a plausible proposal of remote linguistic relationship, one must attempt to eliminate all other possible explanations, leaving a shared common ancestor as the most likely. Greenberg made no attempt to eliminate these other explanations, and the similarities he amassed appear to be due mostly to accident and a combination of these other factors: "I find no evidence whatsoever that [Greenberg's] putative cognate sets ... represent anything other than chance similarities" (Ringe, 1996: 152). In various instances, Greenberg compared arbitrary segments of words, equated words with very different meanings (for example, 'excrement/night/grass'), misidentified many languages, failed to analyze the morphology of some words and falsely analyzed that of others, neglected regular sound correspondences, failed to eliminate loanwords, and misinterpreted well-established

findings. The Amerind 'etymologies' proposed are often limited to a very few languages of the many involved. (For details and examples, see Adelaar, 1989; Berman, 1992; Campbell, 1988, 1997a; Kimball, 1993; McMahon and McMahon, 1995; Poser, 1992; Rankin, 1992; Ringe, 1992, 1996). Finnish, Japanese, Basque, and other randomly chosen languages fit Greenberg's Amerind data as well as or better than any of the American Indian languages do. Greenberg's method has proven incapable of distinguishing implausible relationships from Amerind generally (Campbell, 1988; Campbell, 1997b).

In short, it is with good reason Amerind has been rejected.

Bibliography

- Adelaar W F H (1989). 'Review of Language in the Americas, by Joseph H. Greenberg.' *Lingua* 78, 249–255.
- Berman H (1992). 'A comment on the Yurok and Kalapuya data in Greenberg's Language in the Americas.' *International Journal of American Linguistics* 58, 230–233.
- Brown C H & Witkowski S R (1979). 'Aspects of the phonological history of Mayan-Zoquean.' *International Journal of American Linguistics* 45, 34–47.
- Campbell L (1988). 'Review of Language in the Americas, by Joseph H. Greenberg.' *Language* 64, 591–615.
- Campbell L (1994). 'The American Indian classification controversy: an insider's view.' *Mother Tongue* 23, 41–55.
- Campbell L (1997a). *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press.
- Campbell L (1997b). 'Genetic classification, typology, areal linguistics, language endangerment, and languages of the north Pacific rim.' In Miyaoka *et al.* (eds.) *Languages of the North Pacific Rim*, vol. 2. Kyoto: Kyoto University. 179–242.
- Campbell L (1998). 'Nostratic: a personal assessment.' In Joseph B & Salmons J (eds.) *Nostratic: sifting the evidence*. Amsterdam: John Benjamins. 107–152.
- Campbell L (1999). 'Nostratic and linguistic palaeontology in methodological perspective.' In Renfrew C & Nettle D (eds.) *Nostratic: evaluating a linguistic macrofamily*. Cambridge: The McDonald Institute for Archaeological Research. 179–230.
- Campbell L (2003). 'How to show languages are related: methods for distant genetic relationship.' In Joseph B & Janda R (eds.) *Handbook of historical linguistics*. Oxford: Blackwell. 262–282.
- Campbell L & Poser W (2008). *Language Classification: History and Method*. Cambridge: Cambridge University Press.
- Dixon R M W (2002). *Australian languages*. Cambridge: Cambridge University Press.
- Dolgopolsky A (1986). 'A probabilistic hypothesis concerning the oldest relationships among the language families in northern Eurasia.' In Shevoroshkin V V & Markey T L (eds.) *Typology, relationship, and time: a*

- collection of papers on language change and relationship by Soviet linguists. Ann Arbor: Karoma Publishers. 27–50.
- Fleming H C (1987). 'Review article: towards a definitive classification of the world's languages (review of A guide to the world's languages, by Merritt Ruhlen).' *Diachronica* 4, 159–223.
- Greenberg J H (1987). *Language in the Americas*. Stanford: Stanford University Press.
- Hoenigswald H M (1990). 'Is the "comparative" method general or family specific?' In Baldi P (ed.) *Linguistic change and reconstruction methodology*. Berlin: Mouton de Gruyter. 375–383.
- Illich-Svitych V M (1990). 'The Nostratic reconstructions of V. Illich-Svitych, translated and arranged by Mark Kaiser.' In Shevoroshkin V (ed.) *Proto-languages and proto-cultures: materials from the first international interdisciplinary symposium on language and prehistory*. Bochum: Brockmeyer. 138–167.
- Jakobson R (1960). 'Why "mama" and "papa"?' In Kaplan B & Wapner S (eds.) *Perspectives in psychological theory*. New York: International Universities Press. 21–29. (Reprinted 1962 *Roman Jakobson selected writings*, vol. 1: *phonological studies*. The Hague: Mouton. 538–545.)
- Janhunen J (1989). 'Any chances for long-range comparisons in North Asia?' *Mother Tongue* 6, 28–30.
- Kaiser M & Shevoroshkin V (1988). 'Nostratic.' *Annual Review of Anthropology* 17, 309–330.
- Kimball G (1992). 'A critique of Muskogean, "Gulf," and Yukian material in Language in the Americas.' *International Journal of American Linguistics* 58, 447–501.
- McMahon A & McMahon R (1995). 'Linguistics, genetics and archaeology: internal and external evidence in the Amerind controversy.' *Transactions of the Philological Society* 93, 125–225.
- Poser W J (1992). 'The Salinan and Yurumanguí data in *Language in the Americas*.' *International Journal of American Linguistics* 58, 202–229.
- Rankin R L (1992). 'Review of *Language in the Americas*, by Joseph Greenberg.' *International Journal of American Linguistics* 58, 324–350.
- Ringe D A Jr (1992). 'On calculating the factor of chance in language comparison.' *Transactions of the American Philosophical Society* 82(1), 1–110.
- Ringe D A Jr (1996). 'The mathematics of "Amerind." ' *Diachronica* 13, 135–154.
- Ruhlen M (1994). *On the origin of languages: studies in linguistic taxonomy*. Stanford: Stanford University Press.
- Shevoroshkin V (1989). 'A symposium on the deep reconstruction of languages and cultures.' In Shevoroshkin V (ed.) *Reconstructing languages and cultures: materials from the first international interdisciplinary symposium on language and prehistory*. Bochum: Brockmeyer. 6–8.
- Teeter K V (1964). 'Algonquian languages and genetic relationship.' In *Proceedings of the 9th International Congress of Linguists*. The Hague: Mouton. 1026–1033.

Louisiana Creole

T Klingler, Tulane University, New Orleans, LA, USA

© 2006 Elsevier Ltd. All rights reserved.

This French-lexifier creole is spoken by an estimated 10 000–20 000 persons (reliable figures are not available) residing mainly in southwestern Louisiana. Most speakers live along or near the Bayou Teche, especially in the parishes of St. Landry, St. Martin, and Lafayette, but there are also pockets of speakers in several other parishes. Although it is commonly associated with African Americans and Creoles of color, Louisiana Creole (LC) is also the first language of many European Americans. The language has long coexisted with regional varieties of French, often referred to collectively as Cajun, and it is at least in part the continued influence of these varieties that explains why LC is structurally less distant from French than are the French-lexifier creoles of the Caribbean. LC shares a number of important features with Haitian Creole (e.g., the progressive marker *ape*, the verb *gen* 'to have,' and the possessive particle *kèl/*

tchèl), and some linguists maintain that LC had as its origin a creole or pre-creole language imported to Louisiana from the French colony of Saint-Domingue before it became the free republic of Haiti. However, evidence that LC's development predated the significant population migration from Saint-Domingue to Louisiana in the early 19th century casts doubt on this claim and strengthens the possibility that LC is indigenous to the region. Today, the future of LC remains uncertain, since most fluent speakers are now elderly and the language is not being passed on to younger generations.

LC varies considerably according to region, ethnic group, and social context. The linguistic situation in Louisiana is often said to form a speech continuum, with the type of LC that is furthest removed from French constituting the basilectal pole, and Cajun or, depending on the model used, Referential French constituting the acrolectal pole. Any given utterance, however, may display a greater or lesser quantity of French-like or Creole-like features, such that it may best be assigned to the broad mesolectal range lying between the two poles of the continuum.

Like the other French-lexifier creoles, LC features definite articles that are postposed to the noun (*tab-la* ‘the table,’ *chyen-ye* ‘the dogs’); a personal pronoun system in which all of the pronouns, regardless of function, are derived from the tonic pronouns of French (1 sg. *mo* < *moi*, 2 sg. *to* < *toi*, 3 sg. *li* < *lui*, etc.); and a verbal system that shows very little inflectional morphology but relies instead on a series of markers placed before the verb to express notions of tense, mood, and aspect. The most important of these are the anterior marker *te*; the progressive marker *ape* (*e* in Pointe Coupee Parish); the future

marker *a*, *va*, or *ale*; and the conditional marker *se*: *Ye te ka lir* ‘They could read’; *Lavach-la ape kòmanse dòn dule* ‘The cow is beginning to give milk’; *Vou pa kwa l a chinen?* ‘Don’t you think he’ll win?’

Bibliography

- Klingler T A (2003). *If I could turn my tongue like that: the creole language of Pointe Coupee Parish, Louisiana*. Baton Rouge: Louisiana State University Press.
Neumann I (1985). *Le créole de Breaux Bridge, Louisiane*. Hamburg: Buske.

Luganda

F Katamba, Lancaster University, Lancaster, UK

© 2006 Elsevier Ltd. All rights reserved.

Location and Genetic Affiliation

Luganda (Ganda), a Bantu language of Uganda, is the mother tongue of 3 015 980 speakers (a little more than 16% of the population of Uganda); with an additional 1 million second language speakers, Luganda is the most widely spoken language in Uganda. It belongs to the Narrow Bantu subgroup of the Bantu sub-branch of the Benue-Congo branch of Niger-Congo. It is classified as Zone J15 in Guthrie’s classification system for Bantu.

Basic Phonology and Orthography

The Luganda orthography is essentially phonemic. The consonants and vowel phonemes are listed in **Tables 1** and **2**, respectively. IPA symbols corresponding to the standard letters are shown in brackets for palatal and velar consonants. As seen in Example (1), gemination (indicated by double letters) is phonemic for both consonants (C) and vowels (V). The typical syllable is

CV or CVV. The only consonant clusters allowed are NC (i.e., nasal + consonant), as in *nkola*, and for consonant plus glide, as in *mukwano* ‘friendship’:

- (1) ogula ‘we buy’ oggula ‘you open’
nkola ‘I work’ nkoola ‘I weed’

Tone is phonemic. However, it is not marked in the standard orthography, and that convention is followed here except when tone is being discussed.

On the surface, there is a contrast between low (L) tone, which is unmarked, high (H) (‘) tone, and falling (HL) (^) tone. There are no rising tones.

- (2) ki sa ni ri zo (L L L L L) ‘comb’
ku wó la (L H L) ‘to become cool’
ku wólá (L H H) ‘to lend money’
bí táâ (L HL) ‘gourds’

Basic Morphology

Luganda has rich, agglutinative morphology. The typical noun has the following structure (PP = pre-prefix; CP = class prefix):

- (3) PP CP ROOT
o- mu- wala ‘girl’

Table 1 Consonants

| Consonant | Bilabial | Labiodental | Alveolar | Palatal | Velar | Labiovelar |
|--------------|----------|-------------|----------------|---------|--------|------------|
| Stops | | | | | | |
| Voiceless | p | | t | c | k | |
| Voiced | b | | d | j [ɟ] | g | |
| Fricatives | | | | | | |
| Voiceless | | f | s | | | |
| Voiced | | v | z | | | |
| Nasals | m | | n | ny [ɲ] | ng [ŋ] | |
| Approximants | | | l ^a | y (j) | | w |

^aThough [l] and [r] are allophones of the phoneme /l/, they are represented by separate letters in the orthography. The letter ‘r’ is used after front vowels and ‘l’ is used elsewhere.

Table 2 Vowels

| Vowel | Front | Central | Back and round |
|-------|-------|---------|----------------|
| High | i | | u |
| Mid | e | | o |
| Low | | a | |

The 21 noun classes, each one of them marked by a prefix, are normally paired for singular and plural, as in Example (4):

| | | | | | |
|---------------|-------|----|-----|------|-----------|
| (4) Singular: | CLASS | PP | CP | STEM | |
| | 1 | o- | mu- | wala | 'girl' |
| | 7 | e- | ki- | bira | 'forest' |
| Plural: | 2 | a- | ba- | wala | 'girls' |
| | 8 | e- | bi- | bira | 'forests' |

The noun class numbering system is standardized for all Bantu languages.

There is concord in a noun phrase between a noun and any dependent adjectives and determiners:

- (5a) o- mu- wala o- no o- mu- nene
 PP CP1 girl CP1 this PP CP1 big
 'this big girl'
- (5b) e- ki- bira ki- no e- ki- nene
 PP CP7 girl CP7 this PP CP7 big
 'this big forest'

As seen in Example (6), verbs, compared to nouns, are morphologically more complex (NEG, negation; SM, subject marker; TM, tense marker; FUT, future; OM, object marker; ES, extension suffix; APPL, applicative; FV, final vowel):

- (6) te- ba- li- ki- n- deet- er- a
 NEG SM TM(FUT) OM1 OM2 bring ES (APPL) FV
 not they future it me bring for FV
 'They will not bring it for me.'

There is also an elaborate tense/aspect system, with distinct negative and positive paradigms:

| (7) Tense/aspect | Positive | Negative |
|------------------|----------|----------|
| Past: | nنالابا | سالابا |
| Near past: | nنالابيه | سالابيه |
| Immediate past: | ندابيه | سيرابيه |
| Present: | ندابا | سيرابا |
| Near future: | ننااسوما | سياسوما |
| General future: | نديرابا | سيرابا |

Basic Syntax

The basic word order in unmarked declarative sentences is subject-verb-object:

- (8a) a- ba- wala ba- a- lab- a e- m- bwa
 PP- CP2- girl CP2- PAST- see- FV PP- CP9- dog
 'The girls saw the dog.'
- (8b) e- m- bwa y- a- lab- a a- ba- wala
 PP- CP9- dog CP9- PAST- see- FV PP- CP2- girl
 'The dog saw the girls.'

Typically, the head precedes its modifier: e.g., in noun phrases, nouns precede determiners and adjectives (see Examples (5a) and (5b)).

Bibliography

- Ashton E O, Muliira E M K, Ndawula E G M & Tucker A N (1954). *A Luganda grammar*. London: Longmans, Green and Co.
- Cole D T (1967). *Some features of Ganda linguistic structure*. Johannesburg: Witwatersrand University Press.
- Hyman L M & Katamba F (1993). 'A new approach to tone in Luganda.' *Language* 6, 34–67.

Luo

G J Dimmendaal, University of Cologne, Cologne, Germany

© 2006 Elsevier Ltd. All rights reserved.

One of the major Nilotic (Nilo-Saharan) languages in terms of number of speakers, Luo (also known as Nilotic Kavirondo), is spoken by approximately 3.5 million people mainly in western Kenya, northern Tanzania, and eastern Uganda. Together with Acholi, Adhola, Alur, Kumam, and Lango, Luo forms the

Southern Lwoo cluster within the Western Nilotic branch of Nilotic. The Luo orthography was developed at the beginning of the 20th century. There is a growing body of literature in this Nilotic language, which is also used in the educational system in Kenya.

Luo is among the few Nilotic languages that has also been studied in detail by native speakers, e.g., Okoth-Okombo (1982) and Omondi (1982). One of the pioneers of African linguistics, Archibald Tucker, also produced a grammar of Luo, published posthumously as Tucker (1994). As shown in these

studies, Luo has a classical two-tone system with downdrift, downstep, as well as upstep. As is common in a wide range of languages ranging from Senegal to Ethiopia, it also has vowel harmony based on the position of the tongue root. Luo appears to have retained relatively few prototypical Nilotic features at the morphosyntactic level, presumably as a result of contact with Niger-Congo languages at different periods in time. One stratum, which seems to have affected all Southern Lwoo languages, appears to be due to contact with Ubanguian (Niger-Congo) languages. Another, more recent stratum resulted from intensive contact between Luo and neighboring Bantu (Niger-Congo) languages (cf. Rottland and Okoth-Okombo, 1986; Dimmendaal, 2001; and Storch, 2003).

One manifestation of the intensive lexical and structural borrowing from Bantu is the development of noun class prefixes in Luo. In addition to borrowed prefixes, one finds prefixes that developed from nominal roots, as in *dhó-lúô* 'the Luo language' (from *dhok* 'mouth'); *jà-lúô/jò-lúô* 'Luo person (sg/pl)' from *jal* (sg), *jol* (pl) 'guest, stranger'.

The common constituent order in Luo is SVO. Other members of the Lwoo cluster, such as Anywa or Pàri, allow for OVS order and they inflect postverbal subjects with (ergative) case. Luo does not have case marking. Consequently, although VS order may be used in Luo to express presentative

focus with intransitive predicates, the postverbal subject is not inflected for case. Compared again to Anywa and Pàri, Luo has a reduced system of verbal derivation, using prepositions instead to modify the valency of verbs. On the other hand, Luo developed tense marking on the verb, parallel to neighboring Bantu languages.

Bibliography

- Dimmendaal G J (2001). 'Language shift and morphological convergence in the Nilotic area.' *Sprache und Geschichte in Afrika* 16/17, 83–124.
- Okoth-Okombo D (1982). *Dholuo morphophonemics in a generative framework. Language and Dialect Atlas of Kenya, Supplement 2*. Berlin: Dietrich Reimer.
- Omondi L N (1982). *The major syntactic structures of Dholuo. Language and Dialect Atlas of Kenya, Supplement 1*. Berlin: Dietrich Reimer.
- Rottland F & Okoth-Okombo D (1986). 'The Suba of Kenya: a case of growing ethnicity with receding language competence.' *Afrikanistische Arbeitspapiere* 7, 115–126.
- Storch A (2003). 'Dynamics of interacting populations: language contact in the Lwoo languages of Bahr El-Ghazal.' *Studies in African Linguistics* 32(1), 63–93.
- Tucker A N with Creider C A (1994). *A grammar of Kenya Luo (Dholuo)*. In Bender M L, Rottland F & Cyffer N (eds.) *Nilo-Saharan Linguistic Analyses and Documentation*, 8. Cologne: Rüdiger Köppe.

Luxembourgish

G Newton, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

Luxembourgish (*Lëtzebuergesch*), genetically related to German, is traditionally grouped with the West Moselle Franconian dialects. However, early Salic Frankish influence and later close attachment to the Low Countries, France, and Spain have allowed it to develop an identity separate from that of the neighboring dialects in Germany. Earliest documents from the area date from the 9th century, with modern literary forms beginning in the 1820s. Various orthographies exist, including the strictly phonemic *Lezebuurjer Ortografi* (1946). Little used, this remained official until replaced in 1975 by the system of the *Luxemburger Wörterbuch*. Modifications to this were introduced in 1999.

In 1939, Luxembourg naturalization was made dependent on knowledge of the language. In 1984, *Lëtzebuergesch* was legally acknowledged as the national language of the Grand Duchy. Syntactically, *Lëtzebuergesch* is similar to German, although case loss has reduced the possibilities of object-verb-subject (OVS) ordering. Parataxis predominates, though hypotaxis is frequent (subject-object-verb (SOV) ordering). In morphology, nominative and accusative have fallen together, assuming accusative form. Third-person pronouns show northern /h/ (NHG = New High German): *hien*, *hatt*, *hinen* NHG *er*, *es*, *ihnen* 'he, it, them'. Noun plurals are most commonly in <e(n)>, though other patterns occur. There is, however, no plural in <s>. The pronouns *mir* NHG *wir* 'we' and *dir* NHG *ihr/Sie* 'you (plural and polite)' arise from false division of verbal endings. NHG *uns* 'us' appears as *äis/eis* (koine) and *ons* (Luxembourg

city). Adjectival comparison is chiefly with *méi* NHG *mehr* 'more', though occasional synthetic forms occur. In compound nouns, a linking <s> is frequently present, e.g., *Plastikstut* NHG *Plastiktasche* 'plastic bag', *Autosdier* NHG *Autotür* 'car door'. Tenses comprise present (*ech gesinn* NHG *ich sehe*), perfect (*ech hu gesinn* NHG *ich habe gesehen*), and pluperfect (*ech hat gesinn* NHG *ich hatte gesehen*); the future can be periphrastic (*ech wäerd gesinn* NHG *ich werde sehen*), but is mainly a function of the present tense (*ech kommen iwwer* NHG *ich komme übermorgen*). Some indicative (*ech gesouch* NHG *ich sah*) and subjunctive (*ech geséich* NHG *ich sähe*) preterites also occur, though these are more frequent in the north (Oesling); pluperfect subjunctives (*ech hätt gesinn* NHG *ich hätte gesehen*) occur frequently. The auxiliary verb *ginn* NHG *geben* = *werden* 'give = become' is used to form analytical conditionals (*ech géif gesinn* NHG *ich würde sehen*) and passives (*ech gouf gesinn* NHG *ich wurde gesehen* 'I was seen'). Present tense first-person singulars inflect <e(n)> (*ech sangen* NHG *ich singe*).

Consonants are in voiced/voiceless opposition, with final neutralization. The High German sound shift is incomplete (*dat* NHG *das/dass* 'that', *op* NHG *auf* 'on', *Pond* NHG *Pfund* 'pound', *Korf* NHG *Korb* 'basket', *Dall* NHG *Tal* 'valley') and intervocalic /g/ is often absent, e.g., *Won*, *Vull* NHG *Wagen*, *Vogel* 'wagon, bird'. Some velarization of dental nasals (*zéng*, *brong* NHG *zehn*, *braun* 'ten, brown') is also present, though this is stronger in the north of Luxembourg, which also has velarized plosives, e.g., *Lekt*, *néck* (koine *Leit*, *net*) NHG *Leute*, *nicht* 'people, not'. Medial and final /s/ combinations are liable to palatization (*Meeschter* NHG *Meister* 'master'), more strongly in the southwest (*Fënschter* NHG *Fenster* 'window'). Final /n/ is 'mobile' (*Eifler Regel*) and is retained only before a following vowel, /h/ or a dental (*den Dag*, but *de Mann* NHG *der/den Tag*, *der/den Mann*), or at juncture. Middle High German (MHG) <î, û, iu> (*îs* 'ice', *triben* 'drive', *hûs* 'house', *liute* 'people', *hiulen* 'howl') appear as /ɛi/ or /ai/ (*Âis* NHG *Eis*, *dreiwen* NHG *treiben*), /a:ʊ/ (*Haus* NHG *Haus*), /ai/ or /aʊ/ (*Leit* NHG *Leute*, *haulen* NHG *heulen*), and <ie, uo, üe> (*brief* 'letter', *fuoz* 'foot', *vüeze* 'feet') appear as /ei/ (*Bréif* NHG *Brief*), /ou/ (*Fouss* NHG *Fuß*), /ei/ (*Féiss* NHG *Füße*). MHG <ei, ou> (*vleisch* 'flesh', *boum* 'tree') have the reflexes /e:/ (*Fleesch* NHG *Fleisch*), /a:/ (*Bam* NHG *Baum*). However, all of these examples may also be subject to allophonic variation, and shortened forms are common. A shift of West Germanic /i/ to /a/ is also a strong characteristic of the language: *Wand* NHG *Wind* 'wind'. Derounding (*Läffel*, *fënnel* NHG *Löffel*, *fünf* 'spoon, five') and

lowering (*domm* NHG *dumm* 'stupid') are also found, as are elements of 'correction' (an abrupt rise and fall of vowel pitch; only vestigially present in Luxembourgish, e.g., *stäif/stEIF* NHG *steif* 'stiff') and 'circumflexion' (a rise and fall of vowel pitch, accompanied by up to three times normal length), e.g., *den Hals* (nom./acc.) NHG *der/den Hals* 'neck' (not circumflexed); *dem Haals* (dat.) NHG *dem Hals(e)* (circumflexed). Another element is the *Schwebelaut* (a lengthening of consonants), which occasionally marks a difference in meaning, e.g., *voll* (short /l/, MHG *vol*) NHG *voll* 'full', *voll* (long /l:/, MHG *volle*) NHG *voll* 'drunk'.

Sample Text

Bei äis goufe vun 1825

[uechtzéng'honnertfënn'efanzwanzeg] bis haut
verschidde Schreifweise gebraucht, déi all hiert
Guddes haten.

/bai ɛ:is 'goufə fun 'uəxtsɛŋ,hənərt,fənəfan'tsvantseç
bis haut fər'ʃidə 'ʃraifvaizə gə'brauxt dei al hi:rt
gudəz 'hɑ:tən/

NHG *Bei uns wurden von*

*achtzehnhundertfünfundzwanzig bis heute
verschiedene Schreibweisen gebraucht, die alle ihr
Gutes hatten.*

'In Luxembourg from 1825 up to today various
spelling systems were used, which all had their
good points'.

Bibliography

- Berg G (1993). *'Mir wëlle bleiwe, wat mir sin': sozio-linguistische und sprachtypologische Betrachtungen über die luxemburgischen Mehrsprachigkeit*. Tübingen: Niemeyer.
- Braun J (1999). *Eis Sprooch richteg schreiwen*. Bartreng: Rapidpress.
- Bruch R (1953). *Grundlegung einer Geschichte des Luxemburgischen*. Luxembourg: P. Linden.
- Bruch R (1954). *Das Luxemburgische im westfränkischen Kreis*. Luxembourg: P. Linden.
- Bruch R (1955). *Précis populaire de grammaire luxembourgeoise*. Luxembourg: P. Linden.
- Derrmann-Loutsch L (ed.) (2003). *Deutsch Luxemburgisches Wörterbuch*. Luxembourg: Saint Paul.
- Gilles P (1999). *Dialektausgleich im Lëtzebuergereschen: zur phonetisch-phonologischen Fokussierung einer Nationalsprache*. Tübingen: Niemeyer.
- Hoffmann F (1979). *Sprachen in Luxemburg: sprachwissenschaftliche und literarhistorische Beschreibung einer Trilingual-Situation*. Luxembourg: Institut Grand-Ducal.

- Keller R E (1961). 'Luxembourgish.' In Keller R E (ed.) *German dialects*. Manchester: Manchester University Press. 248–298.
- Luxemburger Wörterbuchkommission (1950–1977). *Luxemburger Wörterbuch* (5 vols). Luxembourg: Linden. [Reissued in 1995 as *Lëtzeburger Dictionär* (2 vols).]
- Newton G (1990). 'Central Franconian.' In Russ C V J (ed.) *The dialects of modern German*. London: Routledge. 136–209.
- Newton G (ed.) (1996). *Luxembourg and Lëtzebuergesch: language and communication at the crossroads of Europe*. Oxford: Clarendon Press.
- Schanen F (2004). *Parlons luxembourgeois*. Paris: L'Harmattan.
- Schmitt L F (ed.) (1963). *Luxemburgischer Sprachatlas*. Marburg: Elwert.

This page intentionally left blank

M

Macedonian

V A Friedman, University of Chicago,
Chicago, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Modern Macedonian (*makedonski* in Macedonian) is a South Slavic language (Slavic, Indo-European). It is not to be confused with Ancient Macedonian, an Indo-European language of uncertain (but not Slavic) affiliation, whose most famous speaker was Alexander the Great. Macedonian is closest to Bulgarian and Serbian.

Macedonian is descended from the dialects of Slavic speakers who settled in the Balkan peninsula during the 6th and 7th centuries C.E. The oldest attested Slavic language, Old Church Slavonic, was based on dialects spoken around Salonica, in what is today Greek Macedonia. As it came to be defined in the 19th century, geographic Macedonia is the region bounded by Mount Olympus, the Pindus range, Mounts Shar and Osogovo, the western Rhodopes, the lower course of the river Mesta (Greek Nestos), and the Aegean Sea. Many languages are spoken in this region, but it is the Slavic dialects to which the glossonym *Macedonian* is applied. The region was part of the Ottoman Empire from the late 15th century until 1912 and was partitioned among Greece, Serbia, and Bulgaria (with a western strip of villages going to Albania) by the Treaty of Bucharest in 1913. The modern Republic of Macedonia, in which Macedonian is the official language, corresponds roughly to the southern part of the territory ceded to Serbia plus the Strumica valley. The population is 2 022 547 (2002 census). Outside the Republic, Macedonian is spoken by ethnic minorities in Albania, Bulgaria, Greece, and Kosovo as well as by émigré communities elsewhere. Greece does not recognize the existence of its ethnic minorities, Bulgaria insists that all Macedonians are really Bulgarians, Albania refused to include questions about language and ethnicity in its last census (2001), and there has not been

an uncontested statistical exercise in Kosovo since 1981, so official figures on Macedonian speakers outside the republic are unavailable; estimates range to 700 000.

History

Modern Macedonian literary activity began in the early 19th century among intellectuals attempting to write their Slavic vernacular instead of Church Slavonic. Two centers of Balkan Slavic literacy arose, one in what is now northeastern Bulgaria, the other in what is now southwestern Macedonia. In the early 19th century, all these intellectuals called their language *Bulgarian*, but a struggle emerged between those who favored northeast Bulgarian dialects and those who favored western Macedonian dialects as the basis for what would become the standard language. Northeast Bulgarian became the basis of standard Bulgarian, and Macedonian intellectuals began to work for a separate Macedonian literary language. The earliest known published statement of a separate Macedonian linguistic identity was by Gjorgji Pulevski 1875, but evidence of the beginnings of separatism can be dated to a letter from the teacher Nikola Filipov of Bansko to the Bulgarian philologist Najden Gerov in 1848 expressing dissatisfaction with the use of eastern Bulgarian in literature and textbooks (Friedman, 2000: 183) and attacks in the Bulgarian-language press of the 1850's on works using Macedonian dialects (Friedman, 2000: 180).

The first coherent plan for a Macedonian standard language was published by Krste Misirkov in 1903. After World War I, Macedonian was treated as a dialect of Serbian in Serbia and of Bulgarian in Bulgaria and was ruthlessly suppressed in Greece. Writers began publishing Macedonian works in Serbian and Bulgarian periodicals, where such pieces were treated as dialect literature, but some linguists outside the Balkans treated Macedonian as a separate language. On August 2, 1944, Macedonian became the official language of what was then the People's Republic of Macedonia. Bulgaria recognized both

the Macedonian language and its own Macedonian minority from 1946 to 1948. From 1948 to the 1960s, some Bulgarian linguists continued to recognize Macedonian as a separate Slavic language. When Macedonia declared independence from Yugoslavia in 1991, Bulgaria immediately recognized the state, but not the nationality or the language. In February 1999, the Bulgarian government officially recognized the Macedonian standard language.

Dialects

Macedonian dialects are divided by a major bundle of isoglosses running from northwest to southeast along the River Vardar, swerving southwest at the confluence of the Vardar and the Crna and continuing down the Crna and into Greece southeast of Florina (Lerin in Macedonian), then bifurcating north of Kastoria (Kostur in Macedonian) so that the remaining Macedonian-speaking villages in Greece and Albania form a transitional zone. The western region is characterized by a relatively homogeneous central area and five groups of peripheral dialects centered on towns around the western periphery. The eastern zone has six dialect groups with no regional center. Standard Macedonian is based on the west-central dialects, with elements from other dialects.

Orthography and Phonology

Macedonian is written in the Cyrillic alphabet, following the principle of one letter per sound, as in Serbian Cyrillic. Macedonian has three distinctive letters – ќ, ђ, ѕ – representing the voiceless and voiced dorsopalatal stops and the voiced dental affricate, respectively. Macedonian Cyrillic ђ is, according to the standard (Koneski, 1967: 115), used to represent clear /l/ before consonants, before back vowels, and word-finally, where it can contrast with velar /l/, e.g., бела [bela] ‘white’ F versus беља [bela] ‘trouble’. The contrast is neutralized before front vowels, where only clear /l/ is prescribed. Some educated speakers pronounce ђ as palatal [ɭ], influenced by the Serbian pronunciation of this letter and the fact that the same reflex occurs in the Skopje town dialect. Standard Macedonian has a five-vowel system (a, e, i, o, u), and most dialects outside the west-central area also have schwa, but of different origins in various regions. There is no letter to represent schwa in Macedonian Cyrillic; when it is necessary to do so, an apostrophe is prescribed. The western Macedonian dialects and the standard are characterized by fixed antepenultimate stress, e.g., *vodéničar* ‘miller’, *vodeníčari* ‘millers’, *vodeničárite* ‘the millers’.

Morphology, Syntax, and Lexicon

Macedonian has masculine, feminine, and neuter genders. Aside from plurals and pronouns, the only remnants of Slavic substantival inflection in Macedonian are the masculine and feminine vocative, which are becoming obsolete; oblique forms for masculine proper names and a few kinship terms and other masculine animates, all facultative; and a quantitative plural for inanimate nouns, which is used only sporadically, except in a few common expressions. Macedonian has a three-way opposition in the postposed definite article – *-t* ‘neutral’, *-v* ‘proximal’, *-n* ‘distal’ – although these meanings can be based on speaker attitude as well as physical distance. The example in (1) is illustrative.

- (1) raki-vče-to ké mu go
brandy-DIM- FUT him.DAT it-ACC
 DEF.NEUT
 dade-š na prijatelj-ov od
give-2.sing.PRES to friend-DEF. from
 MASC.PX
 naš-a-na vo frizer-ov
our-FEM.FEM. in freezer-DEF.
 DEF.DS MASC.PX
 ‘Give the little [glass of] brandy to our friend here,
 from that [brandy] of ours, in the freezer here.’

The article attaches to the end of the first nominal in the noun phrase, i.e., not adverbs:

- (2) ne mnogu po-star-i-te deca
not much COMP-old- children
PL-DEFPL
 ‘the children that are not much older’
- edna od mnogu-te naš-i zadač-i
one from many-DEFPL our-PL problems-PL
 ‘one of our many problems’

The Macedonian verb has both aorist/imperfect and perfective/imperfective aspectual oppositions, but imperfective aorists are now obsolete. Perfective presents and imperfects occur only after one of eight modal particles, although perfective presents can also be used in negative questions. Macedonian also developed a new perfect series using the auxiliary *ima* 'have' and an invariant neuter verbal adjective. The synthetic pasts are marked for speaker confirmation, while the descendent of the Common Slavic perfect, using the old resultative participle in *-l* (no longer a true participle, since it cannot be used attributively), is not marked for speaker confirmation and is therefore used when the speaker cannot or will not vouch for the truth of the statement, e.g., because it was reported: *Toj beše vo Moskva* 'He was in Moscow' (I saw him or accept the fact as established). *Toj bil vo Moskva* 'He was in Moscow' (I heard it but was not

there myself, do not vouch for it, or do not believe it [nuance depending on context]). The verbal *l*-form is also used in the inherited Slavic pluperfect (with the auxiliary 'be' in the imperfect) and the inherited conditional (after invariant modal particle *bi*). The new pluperfect is formed with the imperfect of 'have' and the neuter verbal adjective. The new conditional uses the invariant future marker *ke* plus the imperfect (perfective or imperfective) of the main verb. The *bi*-conditional tends to be used for hypothetical apoduses and the *ke* conditional for irrealis.

The following are distinctively Macedonian lexical items: *saka* 'want, like, love', *bara* 'seek', *zboruva* 'speak', *zbor* 'word', *deka* 'that (relativizer)', *vaka* 'in this manner', *olku* 'this many'.

Bibliography

- Apostolski M *et al.* (1969). *Istorija na makedonskiot narod II*. [History of the Macedonian people.] Skopje: Institut za Narodna Istorija.
- Friedman V A (2000). 'The Modern Macedonian standard language and its relation to Modern Macedonian identity.' In Roudometoff V (ed.) *The Macedonian question: culture, historiography, politics*. Boulder, CO: East European Monographs. 173–206.
- Friedman V A (2002). *Macedonian*. (*Languages of the world/materials*, 117). Munich: LinCom Europa.
- Koneski B (1967). *Gramatika na makedonskiot literaturni jazik*. [Grammar of the Macedonian literary language.] Skopje: Kultura.
- Koneski B (1983). *Macedonian historical phonology* (Shevelov G [ed.] *Historical Phonology of the Slavic languages*, Bd. 12). Heidelberg: Carl Winter.
- Kramer C (1986). *Analytic modality in Macedonian*. (Sagner O [ed.] *Slavistische Beiträge* Bd. 198). München: Otto Sagner.
- Kramer C (2003). *Macedonian: A course for beginning and intermediate students*. Madison: University of Wisconsin.
- Lekov I (1968). *Kratka sravnitelno-istoričeska i topologičeska gramatika na slavjanskite ezici*. [A short comparative-historical and typological grammar of the Slavic languages.] Sofia: Bulgarian Academy of Sciences.
- Lunt H (1952). *A grammar of the Macedonian literary language*. Skopje: Državno knigoizdatelstvo.
- Vidoeski B (2004). *The Dialects of Macedonian*. Bloomington, IN: Slavica.

Macro-Jê

E R Ribeiro, Museu Antropológico, Universidade Federal de Goiás, Goiás, Brazil

© 2006 Elsevier Ltd. All rights reserved.

The Macro-Jê stock comprises the Jê family and a number of possibly related language families, all of which are located in Brazil. Macro-Jê is arguably one of the lesser-known language groups of South America, its very existence as a genetic unit being still "a working hypothesis" (Rodrigues, 1999: 165). According to Rodrigues (1986, 1999), whose classification is the most widely accepted among researchers working on Brazilian languages, the 'Macro-Jê hypothesis' comprises 12 different language families: Jê, Kamakã, Maxakalí, Krenák, Purí, Karirí, Yatê, Karajá, Ofayê, Boróro, Guató, and Rikbaktsa. The existence of Jê as a language family has been recognized since early classifications of South American languages (Martius, 1867). 'Jê' is a Portuguese spelling for a Northern Jê collective morpheme ([jɛ] in Apinajé, for instance) that occurs in the names of several Jê-speaking peoples. The term 'Macro-Jê' was coined by Mason (1950), replacing earlier labels, such as 'Tapuya' and 'Tapuya-Jê.'

Comparative Evidence

Recent classifications (Rodrigues, 1986; Greenberg, 1987; Kaufman, 1994) differ as to the precise scope of Macro-Jê, although there is agreement on the inclusion of most of the families (Table 1). Except for Karirí (included only by Rodrigues), Greenberg and Kaufman included all the families listed above. In addition, Greenberg included Chiquitano (also included by Kaufman), Jabutí, and Otí. Given the lack of comprehensive comparative studies, the Macro-Jê status of some of these families is still an open question. Although Guató is included in the stock by all of the aforementioned classifications, a case for its inclusion has yet to be made, beyond the superficial, inconclusive evidence presented so far (Rodrigues, 1986, 1999). On the other hand, a preliminary comparison has revealed compelling evidence for the inclusion of the Jabutí family into the Macro-Jê stock (Voort and Ribeiro, 2004), thus corroborating a hypothesis suggested in the 1930s by ethnographer Curt Nimuendaju (Nimuendaju, 2000: 219–221). Greenberg's main piece of evidence for the inclusion of Chiquitano was the entire set of singular personal prefixes (Greenberg, 1987: 44), which are strikingly similar to the ones found in several Macro-Jê families;

Table 1 The Macro-Jê Hypothesis^a

| | |
|--------------------------|--|
| 1. Jê | †Jeikó <i>Northern Jê</i> : Panará, Suyá, Kayapó, Timbira (Parkatêjê, Pykobjê, etc.), Apinajê <i>Central Jê</i> : Xavánte, Xerênte, †Akroá-Mirim, †Xakriabá <i>Southern Jê</i> : Kaingáng, Xoklêng, †Ingain |
| 2. Kamakã | †Kamakã, †Mongoyó, †Meniën, †Kotoxó, †Masakará |
| 3. Maxakalí | Maxakalí, †Pataxó, †Kapoó, †Monoxó, †Makoni, †Malalí |
| 4. Krenák | Krenák (Botocudo, Borúm) |
| 5. Purí (Coroado) | †Coroado, †Purí, †Koropó |
| 6. Ofayé | Ofayé |
| 7. Rikbaktsá | Rikbaktsá |
| 8. Boróro | Boróro, †Umutína, †Otúke |
| 9. Karajá | Karajá (including four dialects, Southern Karajá, Northern Karajá, Javaé, and Xambioá) |
| 10. Karirí | †Kipeá, †Dzubukuá, †Pedra Branca, †Sabuyá (included by Rodrigues but not Greenberg or Kaufman) |
| 11. Jabutí | Djeoromitxi (Jabutí) Arikapú (included by Greenberg but not Rodrigues or Kaufman) |
| 12. Yatê | Yatê |
| 13. Guató | Guató |
| 14. Chiquitano | Chiquitano (Besiro) (included by Greenberg and Kaufman, but not Rodrigues) |
| 15. Otí | †Otí (Eo-Xavánte) (the inclusion of Otí, proposed only by Greenberg, is not substantiated by the available data) |

^aExtinct languages are indicated by †. Based on Greenberg, 1987; Rodrigues, 1986, 1999; Kaufman, 1994.

convincing lexical evidence, however, has not been presented thus far. As for Otí, a poorly documented language once spoken in southern Brazil, the meager available data do not support its inclusion in the Macro-Jê stock.

The only family-level reconstruction available is Davis (1966), for Proto-Jê. So far, lexical comparative evidence supporting the inclusion of individual families in the Macro-Jê stock has been presented for Kamakã (Loukotka, 1932), Maxakalí (Loukotka, 1931, 1939; Davis, 1968), Purí (Loukotka, 1937), Boróro (Guérios, 1939), Krenák (Loukotka, 1955; Seki, 2002), Karajá (Davis, 1968), Ofayé (Gudschinsky, 1971), Rikbaktsá (Boswood, 1973), and Jabutí

(Voort and Ribeiro, 2004). In addition, some studies have shown very suggestive cases of morphological idiosyncrasies shared by Jê, Boróro, Maxakalí, Karirí, Karajá, and Ofayé (Rodrigues, 1992, 2000b). Thus, although the inclusion of many of the families into the Macro-Jê stock is being further corroborated by additional research, for others (namely Guató, Chiquitano, and Yatê) the hypothesis has yet to be systematically tested. The precise relationship among the suggested members of the stock also remains to be worked out.

Long-Range Affiliations

Greenberg (1987) suggested that Macro-Jê would be related to his Macro-Pano and Macro-Carib stocks, as part of a Jê-Pano-Carib branch of ‘Amerind.’ However, as Rodrigues (2000a) pointed out, Greenberg’s purported evidence does not withstand careful examination. Rodrigues (1985, 2000a) proposed instead a relationship between Tupí, Carib, and Macro-Jê, noting grammatical and lexical similarities among the three language groups (especially between Carib and Tupí). Davis (1968) also mentioned a few lexical similarities between Proto-Jê and Proto-Tupí. Although the evidence presented so far suggests that Rodrigues’s proposal is more plausible than Greenberg’s, any hypothesis of distant genetic relationship at such a level must be considered with caution. Considering that the precise boundaries of Macro-Jê are still uncertain, much more research at the family and stock levels needs to be conducted before such long-range classifications can be proposed on solid scientific grounds.

Location

All Macro-Jê languages are spoken in Brazilian territory, although in the past Otúke (Boróro) and Ingain (Southern Jê), both now extinct, were spoken in Bolivia and Argentina, respectively. Chiquitano, listed as a Macro-Jê language by Greenberg (1987) and Kaufman (1994), is also spoken in Bolivia, as well as in Mato Grosso, Brazil. Although the Jabutí languages and Rikbaktsá are spoken in the southern fringes of the Amazon (Rondônia and northern Mato Grosso, respectively), the overall distribution of Macro-Jê languages is typically non-Amazonian. Yatê, Krenák, and Maxakalí languages are spoken in eastern Brazil, the same having been the case of Purí, Kamakã, and Karirí (all now extinct). Central and Northern Jê tribes, as well as the Boróro and the Ofayé, traditionally occupy the savanna areas of central Brazil. The southernmost Macro-Jê languages are

those belonging to the southern branch of the Jê family, spreading from São Paulo to Rio Grande do Sul. Karajá is spoken along the Araguaia River, in central Brazil. The traditional Guató territory is the Paraguay River, near the Bolivian border. Since several purported Macro-Jê languages were spoken in eastern Brazil, a number of them became extinct early on, under the impact of European colonization. Yatê is a remarkable exception, being the only surviving indigenous language in the Brazilian northeast.

Whereas Guató, Rikbaktsá, Karajá, Krenák, and Ofayé are all single-member families (Table 1), the Jê family has a relatively large number of members, for most of which a fair amount of descriptive material is now becoming available (mostly as graduate theses and dissertations in Brazilian universities). Ofayé has around a dozen speakers, although it is mistakenly listed as extinct by some sources (including earlier editions of *Ethnologue*). Boróro and Maxakalí are the only surviving languages of their respective families. All the languages of the Kamakã, Purí, and Karirí families are now extinct. While documentation on Kamakã and Purí languages consists only of brief wordlists, the Karirí languages Kipeá and Dzubukuá were documented in catechisms (Mamiani, 1698; Bernardo de Nantes, 1709; respectively) and, for Kipeá, a grammar (Mamiani, 1699) – the only published grammar of a non-Tupí language from colonial Brazil. Thus, among the extinct Macro-Jê families, Karirí is the only one for which detailed grammatical information is available. Many of the languages included in the Macro-Jê stock are seriously endangered (Guató, Ofayé, Krenák, and Arikapú are especially so).

Characteristics

When compared with languages of other lowland South American families (such as Carib and Tupí-Guaraní), Macro-Jê languages typically present larger vowel inventories. For instance, Davis (1966) reconstructed, for Proto-Jê, a system of nine oral and six nasal vowels, as well as 11 consonants. Syllabic patterns are rather simple, obstruent clusters being uncommon. Stress is generally predictable. Phonologically contrastive tone oppositions occur in Yatê and Guató (Palácio, 2004). Processes such as nasal spreading and vowel harmony are generally absent. An exception is Karajá, which presents advanced tongue root vowel harmony, a rare phenomenon among South American languages (Ribeiro, 2002a). Another remarkable feature of Karajá is the existence of systematic differences between male and female speech. Female speech is more conservative, male

Table 2 Female versus male speech distinctions in Karajá

| Female speech | Male speech | |
|---------------|--------------|---|
| <i>koworu</i> | <i>oworu</i> | 'wood' |
| <i>dikarê</i> | <i>diarê</i> | 'I' |
| <i>kohã</i> | <i>ohã</i> | 'armadillo' |
| <i>kədãra</i> | <i>ədãra</i> | 'sand' |
| <i>ruku</i> | <i>ru</i> | 'night' |
| <i>beraku</i> | <i>bero</i> | 'river' |
| <i>dəki</i> | <i>dii</i> | '3rd person pronoun' |
| <i>kəbərə</i> | <i>ōbərə</i> | 'to buy' (from Portuguese <i>comprar</i>) |
| <i>kabɛ</i> | <i>abɛ</i> | 'coffee' (from Portuguese <i>café</i>) |
| <i>bəkawa</i> | <i>bəawa</i> | 'firearm' (from Língua Geral <i>mokãáwa</i>) |

speech being characterized, in general, by the deletion of a velar stop occurring in the corresponding female speech form (as a result of consonant deletion, vowel assimilation and fusion may also occur). This is a very productive process, applying even to loanwords (Table 2).

Most Macro-Jê languages have a relatively simple morphology. In most languages (including those of the Jabutí, Karirí, Krenák, Jê, Ofayé, and Maxakalí families), productive inflectional morphology is limited to person marking, the same paradigms being generally shared by nouns, verbs, and adpositions alike. Tense and aspect distinctions are generally conveyed by particles and auxiliaries rather than by inflections (with few apparent exceptions, such as Yatê; cf. Costa, 2004). Noun incorporation is rare, having been reported for a few Northern Jê languages, such as Panará (which also presents postposition incorporation; cf. Dourado, 2002).

In languages with a more robust morphology, such as Karajá, Guató, and Yatê, inflectional morphology tends to be more complex with verbs than with nouns. In Karajá, for example, the verb form includes subject-agreement, voice (transitive, passive, and antipassive), and directional markers ('thither' versus 'hither'), which can be used with evidential purposes (Ribeiro, 2002b); on the other hand, the only category for which nouns inflect is possession (as in most Macro-Jê languages).

The majority of the purported Macro-Jê languages are verb final, with postpositions instead of prepositions and possessor-possessed order in genitive constructions (the exceptions being Guató, Chiquitano, and Karirí). Macro-Jê languages seemingly lack the adjective as an independent part of speech, with adjectival meanings being expressed by nouns or descriptive verbs. Oliveira (2003) offered an in-depth discussion of the properties displayed by 'descriptives' in a particular Macro-Jê language, Apinajé,

illustrating well the issues involved in determining part-of-speech membership in languages in which most inflectional properties tend to be shared by nouns, verbs, and adpositions. In attributive constructions, descriptives follow the word they modify.

Languages such as Maxakalí, Karirí, and Panará are described as being predominantly ergative. In addition, a number of Jê languages are described as presenting an ergative split of some sort. That is the case of Xoklêng (Urban, 1985) and Northern Jê languages such as Kayapó (Silva and Salanova, 2000) and Apinajé. Among the latter, however, ergativity seems to be rather epiphenomenal, being found only in constructions involving nominalized verbs (such as relative clauses; cf. Oliveira, 2003). Syntactic ergativity is rarely found in Macro-Jê, with the exception of Karirí, in which all grammatical criteria (verb inflection, relativization, switch-reference, word order) point to the absolutive argument (S/O) as being the syntactic pivot (Larsen, 1984).

Further Reading

For information on the main literature on Macro-Jê languages, including an overview of their phonological and grammatical characteristics and a short list of possible Macro-Jê cognate sets, see Rodrigues (1999). Proceedings of recent conferences (the 'Encontros Macro-Jê,' which have been taking place periodically since 2001) help to provide an updated picture of Macro-Jê scholarship; the proceedings of the first two meetings were published as Santos and Pontes (2002) and D'Angelis (2004), respectively. Population figures for all Macro-Jê groups (including those now monolingual in Portuguese) can be found in Ricardo (2001).

Bibliography

- Bernardo de Nantes R P Fr. (1709). *Katecismo indico da lingua Kariris*. Lisbon: Valentim da Costa. [Facsimile reproduction by Platzmann J (ed.). *Catecismo da lingua Kariris*. Leipzig: Teubner, 1896.]
- Boswood J (1973). 'Evidências para a inclusão do Aripaktsá no filo Macro-Jê.' In Bridgeman L I (ed.) *Série Lingüística*, 1. Brasília: SIL. 67–78.
- Costa J da (2004). 'Morfologia do verbo em Yaathe.' In D'Angelis (org.) 149–161.
- D'Angelis W da R (org.) (2004). *LIAMES (Línguas Indígenas Americanas) 4. Special Issue, Proceedings of the 2nd 'Encontro de Pesquisadores de Línguas Jê e Macro-Jê,' Campinas, May 2002*.
- Davis I (1966). 'Comparative Jê phonology.' *Estudos Lingüísticos: Revista Brasileira de Lingüística Teórica e Aplicada* 1, 2.10–24.
- Davis I (1968). 'Some Macro-Jê relationships.' *International Journal of American Linguistics* 34, 42–47.
- Dourado L (2002). 'Construções aplicativas em Panará.' *DELTA* 18, 203–231.
- Greenberg J H (1987). *Languages in the Americas*. Stanford: Stanford University Press.
- Gudschinsky S C (1971). 'Ofayé-Xavante, a Jê language.' In Gudschinsky S C (ed.) *Estudos sobre línguas e culturas indígenas*. Brasília: SIL. 1–16.
- Guérios R F M (1939). 'O nexó lingüístico Bororo-Merrime-Caiapó.' *Revista do Círculo de Estudos Bandeirantes* 2, 61–74.
- Kaufman T (1994). 'The native languages of South America.' In Asher R E & Moseley C (eds.) *Atlas of the world's languages*. London: Routledge. 46–76; 14–25 (maps).
- Larsen T W (1984). 'Case marking and subjecthood in Kipeá, Kirirí.' *Berkeley Linguistic Society Proceedings* 10, 189–205.
- Loukotka C (1931). 'La familia lingüística Mašakali.' *Revista del Instituto de Etnología de la Universidad Nacional de Tucumán* 2, 21–47.
- Loukotka C (1932). 'La familia lingüística Kamakan del Brasil.' *Revista del Instituto de Etnología de la Universidad Nacional de Tucumán* 2, 493–524.
- Loukotka C (1937). 'La familia lingüística Coroadó.' *Journal de la Société des Américanistes de Paris* n.s. 29, 157–214.
- Loukotka C (1939). 'A língua dos Patachos.' *Revista do Arquivo Municipal de São Paulo* 55, 5–15.
- Loukotka C (1955). 'Les indiens Botocudo et leur langue.' *Lingua Posnaniensis* 5, 112–135.
- Mamiani L V (1698). *Catecismo da doutrina Christã na lingua brasilica da nação Kiriri*. Lisbon: Miguel Deslandes. [Facsimile: Garcia R (ed.). *Catecismo Kiriri*. Rio de Janeiro: Imprensa Nacional, 1942.]
- Mamiani L V (1699). *Arte de grammatica da lingua brasilica da nação Kiriri*. Lisbon: Miguel Deslandes. [2nd edn.: Rio de Janeiro: Bibliotheca Nacional, 1877.]
- Martius K F P von (1867). *Beiträge zur Ethnographie und Sprachenkunde Amerikas zumal Brasiliens* (2 vols). Leipzig: Friedrich Fleischer.
- Mason J A (1950). 'The languages of South American Indians.' In Steward J H (ed.) *Handbook of South American Indians*, vol. 6. Washington, DC: Smithsonian Institution. 157–317.
- Nimuendaju C (2000). *Cartas do Sertão, de Curt Nimuendajú para Carlos Estevão de Oliveira*. Lisbon: Museu Nacional de Etnologia/Assírio & Alvim.
- Oliveira C C de (2003). 'Lexical categories and the status of descriptives in Apinajé.' *International Journal of American Linguistics* 69, 243–274.
- Palácio A (2004). 'Alguns aspectos da língua Guató.' In D'Angelis (org.). 163–170.
- Ribeiro E R (2002a). 'Directionality in vowel harmony: the case of Karajá.' *Berkeley Linguistics Society Proceedings* 28, 475–485.
- Ribeiro E R (2002b). 'Direction in Karajá.' In Fernández Z E & Ciscomani R M O (eds.) *Memorias, VI Encontro de Lingüística en el Noroeste*, vol. 3. Hermosillo: Editorial UniSon. 39–58.

- Ricardo C A (ed.) (2001). *Povos indígenas no Brasil: 1996/2000*. São Paulo: Instituto Socioambiental.
- Rodrigues A (1985). 'Evidence for Tupí-Carib relationships.' In Klein H E & Stark L R (eds.) *South American Indian languages: retrospect and prospect*. Austin: University of Texas Press. 371–404.
- Rodrigues A (1986). *Línguas brasileiras: para o conhecimento das línguas indígenas*. São Paulo: Edições Loyola.
- Rodrigues A (1992). 'Um marcador Macro-Jê de posse alienável.' In *Anais da 44a. Reunião Anual da SBPC*. São Paulo: SBPC. 386.
- Rodrigues A (1999). 'Macro-Jê.' In Dixon R M W & Aikhenvald A (eds.) *The Amazonian languages*. Cambridge: Cambridge University Press. 164–206.
- Rodrigues A (2000a). 'Ge-Pano-Carib versus Jê-Tupí-Karib: sobre relaciones lingüísticas prehistóricas en Sudamérica.' In Miranda L (ed.) *Actas del I Congreso de Lenguas Indígenas de Sudamérica* (tomo I). Lima: Universidad Ricardo Palma. 95–104.
- Rodrigues A (2000b). 'Flexão relacional no tronco Macro-Jê.' *Boletim da Associação Brasileira de Linguística* 25, 219–231.
- Santos L dos & Pontes I (eds.) (2002). *Línguas Jê: estudos vários*. Londrina: Editora UEL.
- Seki L (2002). 'O Krenak (Botocudo/Borum) e as línguas Jê.' In Santos & Pontes (eds.). 15–40.
- Silva M A R & Salanova A (2000). Verbo y ergatividad escindida en Mebêngôkre. In Voort H van der & Kerke S van der (eds.) *Indigenous Languages of Lowland South America*. Leiden: University of Leiden. 225–242.
- Urban G (1985). 'Ergativity and accusativity in Shokleng (Gê).' *International Journal of American Linguistics* 51, 164–187.
- Voort H van der & Ribeiro E R (2004). 'The westernmost branch of Macro-Jê.' Paper presented at the 2nd Workshop 'Linguistic Prehistory in South America,' September 4–10. University of Oregon, Department of Linguistics.

Madang Languages

A Pawley, Australian National University, Canberra, ACT, Australia

© 2006 Elsevier Ltd. All rights reserved.

The Madang group, containing about 100 languages, is the largest well-defined branch of the Trans New Guinea (TNG) family, which dominates the large island of New Guinea (see **Trans New Guinea Languages**). The Madang subgroup occupies the central two-thirds of Madang Province in north central Papua New Guinea (see **Figure 1**). In the east the group's immediate neighbors are languages of the Finisterre–Huon branch of TNG. In the high mountain valleys to the south lie the Goroka, Chimbu–Wahgi, and Engan branches and to the west are unrelated languages, members of the Lower Sepik–Ramu family. The most important innovations defining the Madang subgroup are the replacement of the Proto-TNG independent pronouns **na* '1sg,' **nga* '2sg,' and **ya* '3sg' by Proto-Madang **ya*, **na* and **nu*, respectively (Ross, 2000).

The Madang group probably broke up more than 5000 years ago, after diverging from its TNG relatives in the central highlands. This rough estimate of time depth is based chiefly on lexicostatistical agreements between languages belonging to different primary branches within Madang, which are of the order of 5–15%, lower than those between the major branches of Indo-European. The whole of Madang Province

has an area smaller than the Netherlands but contains some 150 languages. Most members of the Madang group have between 500 and 2000 speakers and none has more than about 20 000. This extreme linguistic fragmentation reflects both the considerable time depth of the Madang subgroup and fact that until the colonial era political units in New Guinea seldom exceeded a few hundred people.

The first written records of Madang languages were made in the 1870s but to this day most of the languages are documented only by word lists and sketchy grammatical notes (Z'graggen, 1975b gives a history of research and Carrington, 1996 contains a near-exhaustive bibliography). The best-documented languages are probably Amele (Roberts, 1981, 1987, 1991), Kalam (Pawley, 1966, 1987, 1993; Lane, 1991; Pawley *et al.*, 2000; Pawley and Bulmer, in press), and Kobon (Davies, 1980, 1981, 1985). There are detailed grammars of several other languages including Anamuxra (a.k.a. Ikundun or Anamgura) (Ingram, in press), Tauya (MacDonald, 1990) and Usan (Reesink, 1987).

Much of the published comparative work on Madang languages is due to John Z'graggen (1971, 1975a, 1975b, 1980a, 1980b, 1980c, 1980d). He posited a 'Madang–Adelbert Range subphylum' of 98 languages which corresponds closely to the Madang group as defined here, except that Kalam and Kobon (wrongly assigned by Z'graggen following Wurm, 1975 to a putative East New Guinea Highlands microphylum) are now included in Madang

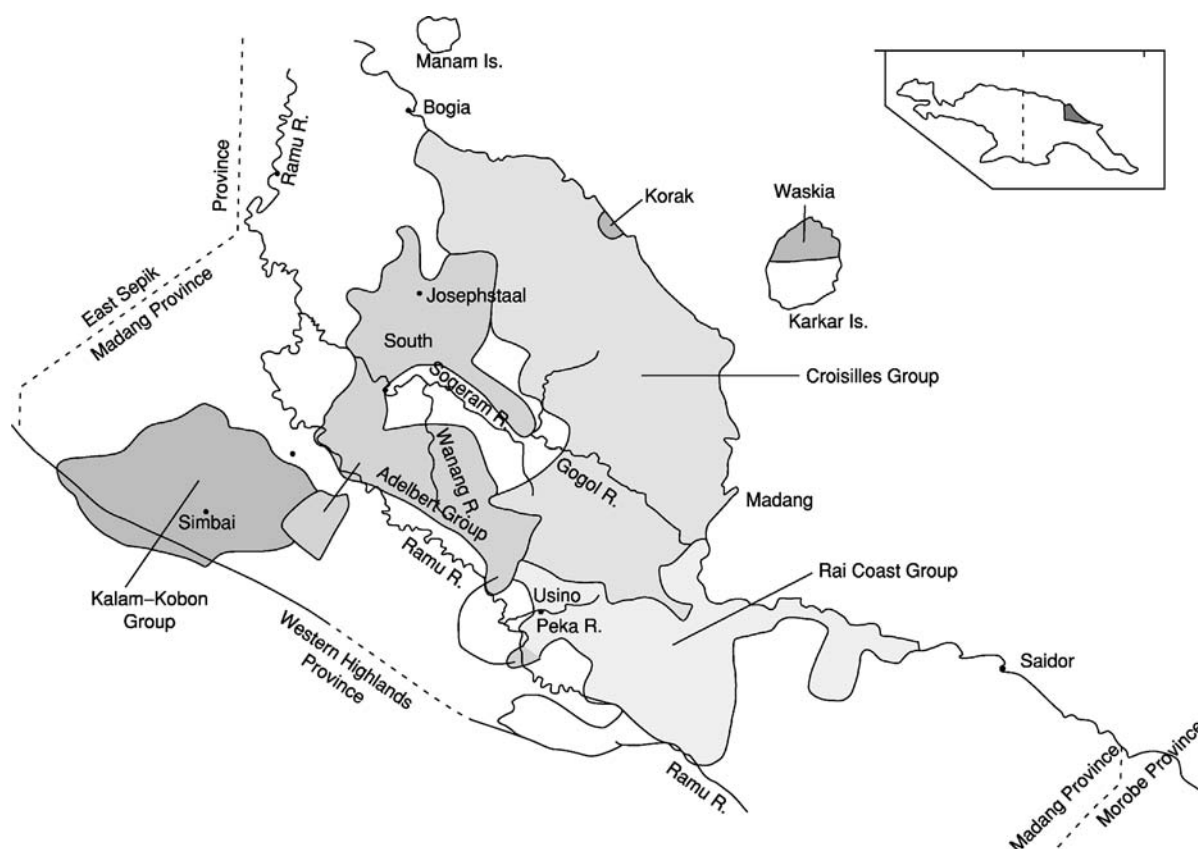


Figure 1 Location of major subgroups of the Madang group.

and Isabi is now excluded (it belongs to the Goroka subgroup of TNG).

Z'graggen also tentatively proposed an internal classification on typological and lexicostatistical grounds. Recent (and largely unpublished) comparative work using more classical subgrouping methods has led to various revisions in the subgrouping (Pawley, 1998; Pawley and Osmond, 1997; Ross, 2000). Five main branches can be distinguished based on innovations in the pronouns (Ross, 2000) and other criteria:

- The Rai Coast group, consisting of about 30 languages, extends along the coastal lowlands from around the mouth of the Gogol River eastwards almost to the mouth of the Mot, and in places extends inland as far south as the Ramu River.
- The Croisilles group of some 50 languages subsumes the 'Mabuso' group and most languages of the 'North Adelbert' group proposed by Z'graggen (1975). Croisilles languages occupy the central Madang coast from the Gogol River north almost as far as Bogia, and cover much of the hinterland west and north of Madang town.
- The South Adelbert group contains 14 languages. Twelve are centered in the South Adelbert Range

north of the Ramu River. The other two, Gants and Faite, are spoken south of the Ramu in separate pockets in or close to the Bismarck Range.

- Waskia and Korak, spoken on Karkar Island and on the coast just west of this, form another group.
- A fifth group consists of Kalam and Kobon (each a chain of diverse dialects), spoken around the junction of the Bismarck and Schrader Ranges where Madang Province meets Western Highlands Province.

Structural Characteristics of Madang Languages

Phonology

A good many Madang languages have syllables of the shape (C)V and (word finally) CVC, five vowels and between 15 and 20 consonants including series of nasals and oral and prenasalized (or voiceless and voiced) obstruents with contrasts at bilabial, apical, and velar (and often palatal) positions. Members of the South Adelbert and Kalam-Kobon groups resemble unrelated languages of the Sepik and Lower Sepik-Ramu families in making heavy use of a high

central or mid central vowel which in some contexts is nonphonemic, being epenthetically inserted between consonants (Biggs, 1963; Pawley, 1966; Ingram, forthcoming).

Grammar

The preferred order of constituents in verbal clauses is SOV but OVS often occurs as a marked structure. Adpositions follow the verb, determiners and possessors follow the noun. Case marking is generally absent or little developed. Most languages organize pronominal affixes to show a nominative–accusative/dative contrast.

Common nouns are an open class but there are several closed classes of nominal roots such as kinship terms and locatives. In Kalam and Kobon, verb roots are a small closed class of about 130 members but in most Madang languages they are more numerous and probably form an open class. Minor word classes include adjectives, adverb roots, and verbal adjuncts (see below). Many Madang languages distinguish singular, dual and plural independent pronouns in three persons. The dual and plural forms are usually distinguished by a suffix.

Morphology is chiefly suffixal. In certain Madang languages, especially in the west, nouns show considerable morphological complexity, including classifying and case-marking suffixes. In others noun morphology is simple but generally kinship nouns take bound possessor pronouns. Sentence-final verbs are typically inflected for tense-aspect-mood and for subject agreement. In some languages transitive verbs also carry a pronominal prefix or proclitic marking object agreement. Dependent verbs in nonfinal clauses are typically marked for relative tense and subject or topic identity with the final verb.

All languages make extensive use of at least one of the following kinds of complex (multiheaded) predicates: (i) in verbal adjunct constructions, a verb, usually carrying a rather general meaning such as ‘make,’ ‘hit,’ or ‘go,’ occurs in partnership with a noninflecting base (the adjunct), which carries more specific meaning; (ii) in serial verb constructions two or more bare verb roots occur in sequence to express a tightly integrated sequence of subevents. Kalam and Kobon allow up to eight or nine verb roots to occur in a single predicate phrase.

In constructions denoting uncontrolled bodily and mental processes (e.g., sweating, sneezing, bleeding, feeling sick) a noun denoting bodily condition is, arguably, the subject. The experiencer is generally marked by an object/dative pronoun and is the direct object.

Long chains of clauses are commonly used to report a sequence of past events that make up a single

episode. Generally, little use of is made of conjunctions to show sequential, conditional, and causal relations. Instead, the main verb in each nonfinal clause carries a suffix which indicates (i) whether the event denoted by the medial verb occurs prior to or simultaneous with that of the final verb, and (ii) whether that verb has the same subject or topic as the next clause. Paragraphlike boundaries are frequently marked by head-to-tail linkage, in which the last clause of the previous sentence is repeated, to begin a new episode.

Bibliography

- Biggs B (1963). ‘A non-phonemic vowel type in Karam, a “pygmy” language of the Schrader Mountains, central New Guinea.’ *Anthropological Linguistics* 5(4), 13–17.
- Carrington L (1996). *A linguistic bibliography of New Guinea*. Canberra: Pacific Linguistics.
- Davies J (1980). *Kobon phonology*. Canberra: Pacific Linguistics.
- Davies J (1981). *Kobon*. Amsterdam: North Holland.
- Davies J (1985). *Kobon dictionary*. MS. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Foley W A (1986). *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Ingram A (in press). *Anamuxra: a language of Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.
- Lane J (1991). *Kalam serial verb constructions*. M.A. thesis, University of Auckland.
- MacDonald L (1990). *A grammar of Tauya*. Berlin: Mouton de Gruyter.
- Pawley A (1966). *The structure of Kalam: a grammar of a New Guinea Highlands language*. Ph.D. thesis, University of Auckland.
- Pawley A (1987). ‘Encoding events in Kalam and English: different logics for reporting experience.’ In Tomlin R (ed.) *Coherence and grounding in discourse*. Amsterdam: John Benjamins. 329–360.
- Pawley A (1993). ‘A language that defies description by ordinary means.’ In Foley W (ed.) *The role of theory in language description*. Berlin: de Gruyter. 87–130.
- Pawley A (1998). *A neogrammarian in New Guinea: searching for sound correspondences in the Middle Ramu*. Print-out. Department of Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Pawley A & Bulmer R (in press). *A dictionary of Kalam with ethnographic notes*. Canberra: Pacific Linguistics.
- Pawley A, Gi S P, Kias J & Majnep I S (2000). ‘Hunger acts on me: the grammar and semantics of bodily and mental processes in Kalam.’ In de Guzman V & Bender B (eds.) *Grammatical analysis in morphology, syntax, and semantics: studies in honor of Stanley Starosta*. Honolulu: University of Hawai’i Press. 153–185.
- Pawley A & Osmond M (1997). *Proto Madang lexical reconstructions*. Research School of Pacific and Asian Studies, Department of Linguistics, Australian National University.

- Reesink G (1987). *Structures and their functions in Usan, a Papuan language of Papua New Guinea*. Amsterdam: John Benjamins.
- Roberts J (1981). *Amele dictionary*. MS. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Roberts J (1987). *Amele*. London: Croom Helm.
- Roberts J (1991). 'A study of the dialects of Amele.' *Language and Linguistics in Melanesia* 22, 67–125.
- Ross M (2000). A preliminary subgrouping of the Madang languages based on pronouns. Printout. Research School of Pacific and Asian Studies, Department of Linguistics, Australian National University.
- Wurm S A (ed.) (1975). *New Guinea area languages 3: Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics.
- Z'graggen J (1971). *Classificatory and typological studies in languages of the Madang district*. Canberra: Pacific Linguistics.
- Z'graggen J (1975a). 'The Madang–Adelbert Range sub-phylum.' In Wurm (ed.) 569–612.
- Z'graggen J (1975b). *The languages of the Madang district, Papua New Guinea*. Canberra: Pacific Linguistics.
- Z'graggen J (1980a). *A comparative word list of the Rai coast languages, Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.
- Z'graggen J (1980b). *A comparative word list of the Northern Adelbert Range languages, Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.
- Z'graggen J (1980c). *A comparative word list of the Mabusu languages, Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.
- Z'graggen J (1980d). *A comparative word list of the Southern Adelbert Range languages, Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.

Madurese

B Nothofer, Universität Frankfurt, Frankfurt, Germany

© 2006 Elsevier Ltd. All rights reserved.

Madurese is the third most widely spoken regional language of Indonesia (after Javanese and Sundanese) and the fourth most widely spoken language in the Austronesian language family (after Malay/Indonesian, Javanese and Sundanese). There are more than 13 million speakers from the island of Madura, from neighboring islands (Kangean Archipelago, Bawean, and Sapudi Islands), and from the northern parts of East Java that were settled by immigrants from infertile Madura. Java has the largest number of Madurese speakers (more than 6 million). Stevens (1968) distinguished two main dialect groups, Maduran and Kangean. Within the Maduran group there are three subgroups: West Madurese, with Bawean and Bangkalan dialects; Central Madurese, with Pamekasan and Sampang dialects; and East Madurese, with Sumenep and Sapudi dialects. The Sumenep dialect is regarded as standard Madurese. The Madurese dialects of East Java vary according to the origins of the speakers. Madurese is a member of the Western Malayo-Polynesian subfamily, which includes the languages of western Indonesia and the Philippines. Lexicostatically, Madurese appears to be most closely related to Malay; it is related to a somewhat lesser degree to the major languages of Java, i.e., Sundanese and Javanese (see Dyen, 1965; Nothofer, 1975). So far, however, no satisfactory qualitative evidence has

been adduced that may contribute to solving the question of the relationship of Madurese to its neighboring languages. Madurese shares with Javanese, Sundanese, Balinese, and Sasak (spoken on Lombok) the existence of speech levels, which serve to indicate the social relationships of the discourse participants. Meanings, for which there are low and high forms, are mostly connected with human beings and refer above all to body parts, body actions, clothing, and personal belongings. Pronouns also have status forms. It is generally assumed that speech levels represent a Javanese innovation and that this system and its higher forms are borrowed from Javanese.

Madurese Phonology

Consonants

The Madurese consonant repertoire resembles that of other western Indonesian languages. However, Madurese has a contrast between a voiceless, voiced, and aspirated stop series. Stevens (1968) characterized these consonant series as follows: (1) voiceless stop, 'voiceless, tense stops', (2) voiced stop, 'voiced, lax stop', and (3) aspirated stop, 'voiceless stop with indifferent tension followed by strong aspiration'. Clynes (1995) suggested that the aspirated stop should rather be described as 'lax voice' or 'whispery voiced'. Only Madurese and Javanese oral stops exhibit five places of articulation, both sharing a phonemic distinction between dental and retroflex (described by Stevens (1968) as 'alveolar stop

Table 1 Consonants in the inherited lexicon of Madurese^a

| | Labial | Dental | Retroflex | Palatal | Velar | Glottal |
|-------------|-----------|-------------|------------|-----------|-----------|----------|
| Stop | | | | | | |
| Voiceless | <i>p</i> | <i>t</i> | <i>t̚</i> | <i>c</i> | <i>k</i> | <i>ʔ</i> |
| Voiced | <i>b</i> | <i>d</i> | <i>d̪</i> | <i>j</i> | <i>g</i> | |
| Aspirated | <i>bʰ</i> | <i>dʰ</i> | <i>d̪ʰ</i> | <i>jʰ</i> | <i>gʰ</i> | |
| Nasal | <i>m</i> | <i>n</i> | | <i>ɲ</i> | <i>ŋ</i> | |
| Fricative | | <i>s</i> | | | | |
| Approximant | <i>w</i> | <i>r, l</i> | | <i>y</i> | | |

^aBased on Clynes (1995) and Davies (1999a).

Table 2 Vowels in the inherited lexicon of Madurese

| | Front | Central | Back |
|------|----------------|----------------|----------------|
| High | /i/ ([i], [e]) | | /u/ ([u], [o]) |
| Mid | | /ə/ ([ɪ], [ə]) | |
| Low | | /a/ ([ɒ], [a]) | |

with larger area of tongue contact than dentals') consonants. Another unusual feature of Madurese is the existence of phonemic consonant gemination. All consonants with the exception of the glottal stop also occur geminated. The consonants occurring in the inherited lexicon of Madurese are shown in **Table 1** (mainly based on Clynes (1995) and Davies (1999a)).

Vowels

Madurese inherited vocabulary has four vowel phonemes, each one having two allophones, which are pairings of high and low vowels (Stevens, 1968; Clynes, 1995; Davies, 1999a). The phoneme /i/ is realized as [i] or [e], /u/ as [u] or [o], /ə/ as [ɪ] or [ə], and /a/ as [ɒ] or [a] (see **Table 2**). In order to account for vowel allophony, Stevens (1968) established the following three categories of Madurese consonants: D_H, voiced and aspirated stops; D_L, voiceless stops, nasals, and intervocalic /s/; and D_N, liquids, glides, /ʔ/, and morpheme-initial and final /s/. The low vowel allophones occur after D_L consonants, in word-initial position, and after immediately preceding low vowels. The high vowel allophones occur following D_H consonants and after immediately preceding high vowels. The D_N consonants do not affect the quality of the vowel. A vowel preceding these consonants determines the quality of the following vowel. If a vowel occurs after a word-initial D_N consonant, this vowel behaves as though it is word-initial.

Madurese vowel harmony results in verb forms with vowels that differ depending on whether they occur in a bare stem or in an active verb in which the initial consonant of the stem is replaced by a homorganic nasal (the prefix N- 'active'). Examples are [mølle]

'active.buy' vs. [billi] 'buy', and [napaʔ] 'active.arrive at' vs. [dɒpɒʔ] 'arrive at' (Davies, 1999a).

Morphology

The major morphological processes are affixation and reduplication (see Stevens, 1968). The verbal affixes include prefixes such as *a-* ('perform action indicated by root; perform action on oneself; to own, have, or use'), *ta-* ('to do unintentionally'), *pa-* ('causative'), and *ka* ('agentless passive'). The prefix N- marks intransitive verbs (with a meaning such as 'agentless action; reflexive action; be like; be in location') and transitive 'active' verbs, whereas 'passive' verbs are marked by *i-*. A verbal circumfix is *ka-an* 'be affected by'. Verbal suffixes are *-a* ('future, conditional, wished for, possible'), *-ag^bi* ('treat like, use object as instrument, perform action with, perform action for, make the object be'), and *-i* ('plural, causative'). Nominal affixes include *pa-* (which derives action nouns from intransitive verbs), *paN-* ('agent, instrument, result of action'), *pa-an* ('location, agent, instrument'), and *-an* ('result of action, that which is affected by action, location of action').

There are three kinds of reduplication: reduplication of the final syllable, total reduplication, and reduplication of the first syllable. The usual meanings with verbs are 'repetition or frequency of action; no specified object or goal'; with nouns the usual meanings are 'plural; groups of objects; instrument used to perform action'.

Writing System

Madurese used to be written in a script derived from Javanese script (*hanacaraka*) that originated from the Pallava script of southern India. Today, Latin orthography is common.

Bibliography

Blust R A (1981). 'The reconstruction of Proto-Malayo-Javanic: an appreciation.' *Bijdragen tot de Taal-, Land- en Volkenkunde* 137, 456–469.

- Clynes A (1995). 'Madurese.' In Tryon D T (ed.) *Comparative Austronesian dictionary. An introduction to Austronesian studies*, (5 vols): part 1: fascicle 1. Berlin, New York: Mouton de Gruyter. 485–494.
- Cohn A C & Lockwood K (1994). 'A phonetic description of Madurese and its phonological implications.' *Working Papers of the Cornell Phonetics Laboratory* 9, 67–92.
- Davies W D (1999a). *Madurese. Languages of the world/materials* 184. München and Newcastle: LINCOM.
- Davies W D (1999b). 'Madurese and Javanese as strict word order languages.' *Oceanic Linguistics* 38, 152–167.
- Dyen I (1965). *A lexicostatistical classification of the Austronesian languages. IUPAL memoir 19; supplement to IJAL* 25. Baltimore: Waverly Press.
- Kiliaan H N (1897). *Madoereesche spraakkunst* (2 vols). Batavia (2nd edition, Semarang: Benjamins, 1911).
- Kiliaan H N (1904). *Madoereesch-Nederlandsch woordenboek*. Leiden: Brill.
- Nothofer B (1975). *The reconstruction of Proto-Malayo-Javanic. Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde* 73. The Hague: Nijhoff.
- Penninga P & Hendriks H (1913). *Practisch Madoereesch-Hollandsch woordenboek*. Semarang: Van Dorp.
- Stevens A M (1965). 'Language levels in Madurese.' *Language* 41, 294–302.
- Stevens A M (1968). *Madurese phonology and morphology. American Oriental Series* 52. New Haven: American Oriental Society.

Malagasy

R Kikusawa, National Museum of Ethnology, Osaka, Japan

© 2006 Elsevier Ltd. All rights reserved.

Geographical Distribution, Dialects, and Speakers

Malagasy is the main language spoken in Madagascar (population approximately 17.5 million according to a 2004 estimate), located off the east coast of Africa. Standard Malagasy, which is based on the Merina dialect spoken in and around Antananarivo, the capital city, is one of the two official languages (along with French) in Madagascar and is used in public contexts and also for education in grade schools and high schools. There are said to be 18 ethnic groups in Madagascar and regional dialects are referred to in association with these groups. Many show phonemic as well as phonetic, lexical, and morphosyntactic features that are different from those in Standard Malagasy. Descriptions are available for some of the dialects (Tsimilaza, 1981; Thomas-Fattier, 1982; Manoro, 1983; Rabenilaina, 1983; Raharinjanahary, 1984; Beaujard, 1998); however, there are a number of others that have not yet been well studied. The Malagasy dialects are considered to form two groups, a western group and an eastern group (Dez, 1963; Gueunier, 1988), distinguished by two sets of regular sound correspondences, Western Malagasy *di* corresponds to Eastern Malagasy *li*, while Western Malagasy *tsi* corresponds

to Eastern Malagasy *ti*. **Figure 1** shows the boundary of the two dialect groups, as well as some regional dialect names.

Genetic Relationships

Malagasy is an Austronesian language. Its most closely related language is considered to be Ma'anyan, a language that belongs to the Barito group of the Western Malayo-Polynesian languages. This implies that the people ancestral to the present Malagasy population migrated from southeast Kalimantan on Borneo Island, where the Barito languages are spoken. This took place probably around 700 A.D., but the exact routes and the reasons for this migration are still not clear (Adelaar, 1989; Dahl, 1991). Some Austronesian features in Malagasy reflect borrowings from genetically related languages, in particular Malay and Javanese, suggesting the possibility of multiple migrations after the initial Austronesian settlement in Madagascar, and/or of continuous contact among the speakers. The language also shows traces of contact with the speakers of such languages as Arabic, Bantu languages (in particular Swahili), and Sanskrit.

Writing Systems

The first writing system introduced to Madagascar was an Arabic script. It was introduced by Muslims in the 12th century, and the people in Taimoro learned it and adapted it to their own phonology

(referred to as *sorabe* ‘great writing/drawing’). Currently, a Latin-based alphabetic system is used, which was introduced in 1823 by an early missionary of the London Missionary Society, David Jones.

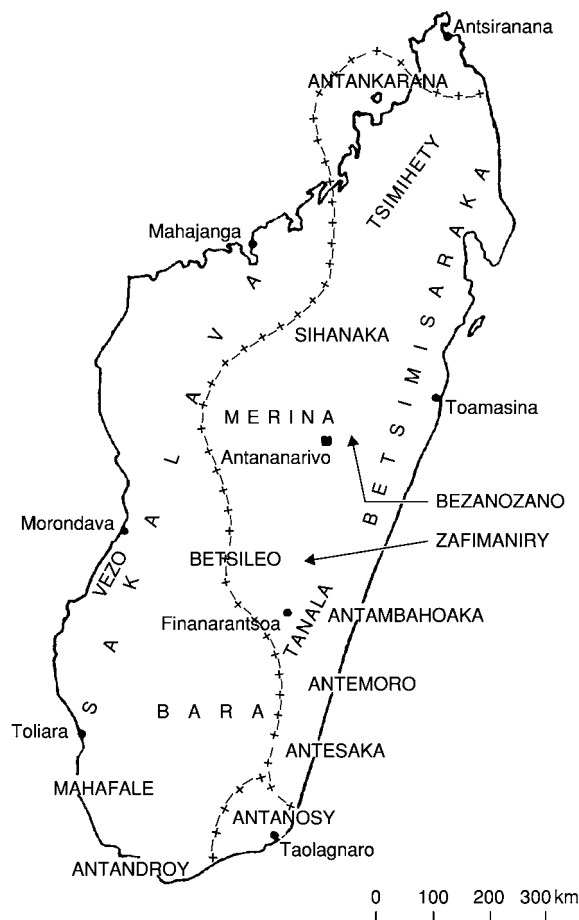


Figure 1 The names of the ethnic groups in Madagascar. The line indicates the boundary between the two dialectal groups, namely Western and Eastern Malagasy.

Linguistic Features of Standard Malagasy

Phonology and Orthography

The letters used in the Malagasy languages and their phonemic properties are shown in Table 1. In Spoken Malagasy, /h/ often disappears, and word-final vowels (sometimes even word-final syllables) become voiceless, or are completely lost.

One of the phonological characteristics of Malagasy is the alternation between spirants and their corresponding stops. This is commonly observed in certain word derivations (such as reduplication and some verb derivations, shown in [1]). Alternation also occurs in compounds when a word starting with a spirant consonant follows the consonant *-n*, such as a genitive marker (2b–2d). It also occurs with transitive verbs when an object is incorporated (2a, 2e). (The symbol ` in examples indicates stress.)

| (1) ROOT | | DERIVED FORM | |
|-----------------|------------------|----------------------|-------------------------------------|
| <i>fito</i> | ‘seven’ | <i>impito</i> | ‘seven times’ |
| <i>sòratra</i> | ‘to write’ | <i>soratsòratra</i> | ‘to write repeatedly’ |
| <i>mihèrika</i> | ‘to look back’ | <i>miherikèrika</i> | ‘to look behind oneself repeatedly’ |
| <i>vàntana</i> | ‘to be straight’ | <i>vantambàntana</i> | ‘to be somewhat straight’ |
| <i>rèraka</i> | ‘to be tired’ | <i>reradrèraka</i> | ‘to be somewhat tired’ |

| | | |
|---|---|---|
| (2a) <i>miàmbina</i> ‘to guard (from)’ + <i>fòdy</i> ‘kind of bird’ | > | <i>miambim-pòdy</i> ‘to guard from birds’ |
|---|---|---|

Table 1 The Malagasy orthography and phonemic system

| Vowels | | | | | | |
|--------------------------------|--------------|--|------------------|----------|------------------|----------|
| ia | | | | | | |
| i, y | [i] | (The letter y is used at the end of a word.) | | | | |
| e | | | | | | |
| o | [u] | | | | | |
| ô | [o] | | | | | |
| Consonants | | | | | | |
| | Nasal | Prenasalized stop | Voiced | | Voiceless | |
| | | | Stop | Spirant | Stop | Spirant |
| Labial | <i>m</i> | <i>mb</i> | <i>b</i> | <i>v</i> | <i>p</i> | <i>f</i> |
| Dentalveolar | <i>n</i> | <i>nd</i> | <i>d</i> | <i>l</i> | <i>t</i> | |
| Alveolar (affricate) | | <i>nj</i> [ndz] | <i>j</i> [dz] | <i>z</i> | <i>ts</i> | <i>s</i> |
| Alveolar trill (or, retroflex) | | <i>ndr</i> [ndr~nr~nd] | <i>dr</i> [dr~d] | <i>r</i> | <i>tr</i> [tr~t] | |
| Velar | <i>ñ</i> [ŋ] | <i>ng</i> [ŋg] | <i>g</i> | | <i>k</i> | <i>h</i> |

The phonetic property is indicated in square brackets, when it is different from the IPA orthography.

- (2b) *òrona* 'nose' + *sàka* 'cat' > *oron-tsàka* 'the nose of a cat'
 (2c) *tràno* 'house' + *hàzo* 'wood' > *tranon-kàzo* 'woodshed'
 (2d) *tràno* 'house' + *vàrotra* 'commerce' > *tranom-bàrotra* 'business association'
 (2e) *lèna* 'wet (with)' + *rano* 'water' > *len-dràno* 'wet with water'

Morphosyntactic Characteristics

Typologically, Malagasy is considered to be a 'pro-drop', verb-initial language that shows the major properties of head-initial languages, with modifiers following the noun and nominal arguments following the verb.

Verbs undergo various morphological derivations that are associated with different sentence structures, as shown here with verbs deriving from the root *pasaka*. In (3), the form *mipasaka* 'to burst open' (the initial consonant appearing as *n* marking the past tense in the example) appears as an intransitive verb requiring only a nominative argument.

- (3) **N-*ipasaka*** ny ovy
 PAST-burst.open DET *potato*
 'the potatoes burst open'

In (4), with the form *manapasaka* 'to smash' (often labeled as 'active voice'), the actor is expressed with a nominative pronoun *aho* 'I,' while in (5) and (6), where the verb forms are *mopasahana* 'to smash something' and *voapasaka* 'have smashed something' (often labeled as 'passive voice'), it is expressed with a genitive pronoun *-ko* 'I (agent).'

- (4) **N-*anapasaka*** ny ovy aho
 PAST-smash DET *potato* *I*
 'I was smashing the potatoes'
 (5) **N-*opasahi-ko*** ny ovy
 PAST-smash-1SG.GEN DET *potato*
 'I smashed the potatoes'
 (6) **Voa-*pasa-ko*** ny ovy
 PERF-smash-1SG.GEN DET *potato*
 'I have finished smashing the
 potatoes / I have inadvertently
 smashed the potatoes.'

The form *manapasahana* 'smash with' (often labeled as 'circumstantial') in (7) typically appears in a relative clause modifying a noun which functions as an instrument or a location.

- (7) **T-amin'** ny sotro no
 PAST-with spoon *that*

- n-*anapasaha-ko*** ny ovy
 PAST-APPL.smash-1SG.GEN DET *potato*
 'it was a spoon with which she was smashing the
 potatoes'

The alternations observed in Malagasy verb morphology, as well as the various sentence structures in which they occur, are of interest in that they correspond both typologically and historically to the 'focus' system in Philippine and Indonesian languages.

Bibliography

- Adelaar K A (1989). 'Malay influence on Malagasy: linguistic and culture-historical implications.' *Oceanic Linguistics* 28(1), 1–46.
 Beaujard P (1998). *Dictionnaire malgache-français: dialecte tañala (sud-est de Madagascar), avec recherches étymologiques*. Paris: Harmattan.
 Dahl O C (1968). *Contes malgaches en dialecte sakalava: textes, traductions, grammaire et lexique*. Oslo: Universitetsforlaget.
 Dahl O C (1991). *Migration from Kalimantan to Madagascar*. Oslo: Norwegian University Press.
 Dahl O C (1993). *Sorabe: revelant l'évolution du dialecte antemoro*. Antananarivo: Trano Printy, Fiangonana Loterana Malagasy.
 Dez J (1963). 'Aperçu pour une dialectologie de la langue malgache.' *Bulletin de Madagascar* 204, 441–451; 205, 507–520; 206, 581–607; 210, 973–994.
 Dyen I (1971). 'Malagasy.' In Sebeok T A (ed.) *Current trends in linguistics 8: Linguistics in Oceania*. The Hague/Paris: Mouton. 211–239.
 Gueunier N-J (1988). 'Dialectologie et lexicostatistique: cas du dialecte malgache de Mayotte (Comores).' *Etudes Océan Indien* 9, 143–170.
 Keenan E L & Polinsky M (1998). 'Malagasy (Austro-nesian).' In Spencer A & Zwicky A M (eds.) *The handbook of morphology*. Oxford/Malden, MA: Blackwell. 563–623.
 Manoro R (1983). *Description morphosyntaxique du Tsimihety (Madagascar)*. Doctoral thesis, University of Paris 7.
 Pearson M & Paul E (eds.) (1996). *The structure of Malagasy 1*. Los Angeles: Department of Linguistics, University of California.
 Rabenilaina Roger-Bruno (1983). *Morpho-syntaxe du malgache: description structurale du dialecte Bàra*. Paris: SELAF.
 Raharinjanahary S (1984). *Aspects de la dialectologie du malgache: morphophonologie: application à l'antanosy*. Doctoral thesis, University of Paris 7.
 Rajaona S (1977). *Problèmes de morphologie malgache*. Fianarantsoa, Madagascar: Ambozontany.
 Rajaonarimanana N (2001). *Grammaire moderne de la langue malgache*. Paris: Langues & Mondes – L'Asiathèque.

- Randriamasimanana C (1986). *The causatives of Malagasy*. Honolulu: University of Hawai'i Press.
- Rasoloson J (2001). *Malagasy-English English-Malagasy dictionary and phrase book*. New York: Hippocrene.
- Thomas-Fattier D (1982). *Le dialecte sakalava du nord ouest de Madagascar*. Paris: SELAF.

- Tsimilaza A (1981). 'Phonologie et morphologie du tsimihety.' Doctoral thesis, University of Nancy 2.
- Vérin P, Kottak C & Gorlin P (1969). 'The glottochronology of Malagasy speech communities.' *Oceanic Linguistics* 8(1), 26–83.

Malay

B Nothofer, University of Frankfurt, Germany

© 2006 Elsevier Ltd. All rights reserved.

Malay, a member of the Malayic language group, belongs to the subfamily of the western Malayo-Polynesian languages of the Austronesian language family. Other Malayic variants that have Proto-Malayic as their common ancestor include Minangkabau, Kerinci, Banjar, Iban, and Jakarta Malay (Adelaar, 1992). Northwestern Borneo is thought to be the homeland of the speakers of the proto-language (Adelaar, 1995; Nothofer, 1997; Collins, 1998). About 2000 years ago some of them migrated to eastern Sumatra, while others remained behind and stayed in northwestern Borneo. Some of the latter traveled south to Ketapang and then crossed over to Bangka and Belitung (Nothofer, 1997). Those remaining in the homeland area are the ancestors of speakers of Malayic Dayak languages (e.g., Iban,

Selako). The Malays who sailed to Sumatra settled the island's east coast. Some moved on into the interior and to the west coast of southern Sumatra. While Middle Malay, Minangkabau, and Kerinci have inland and west coast variants as their origin, Malay itself developed from isolects spoken on the east coast. Later, Malay speakers from the southeast coast of Sumatra established Malay colonies in the Malay Peninsula. Other Malays returned to west Borneo, where they settled the coastal and riverine areas. The isolects spoken by these relatively recent migrants differ considerably from the isolects of Malays who never left Borneo. Coastal Borneo has other Malay isolects such as Sarawak Malay, Brunei Malay, Kutai Malay, and Banjar, perhaps as a result of a clockwise settlement that originated in western Borneo (**Figure 1**).

Malay, the native language of the powerful kingdoms along the shores of the Straits of Malacca through which all traders from the west and the east

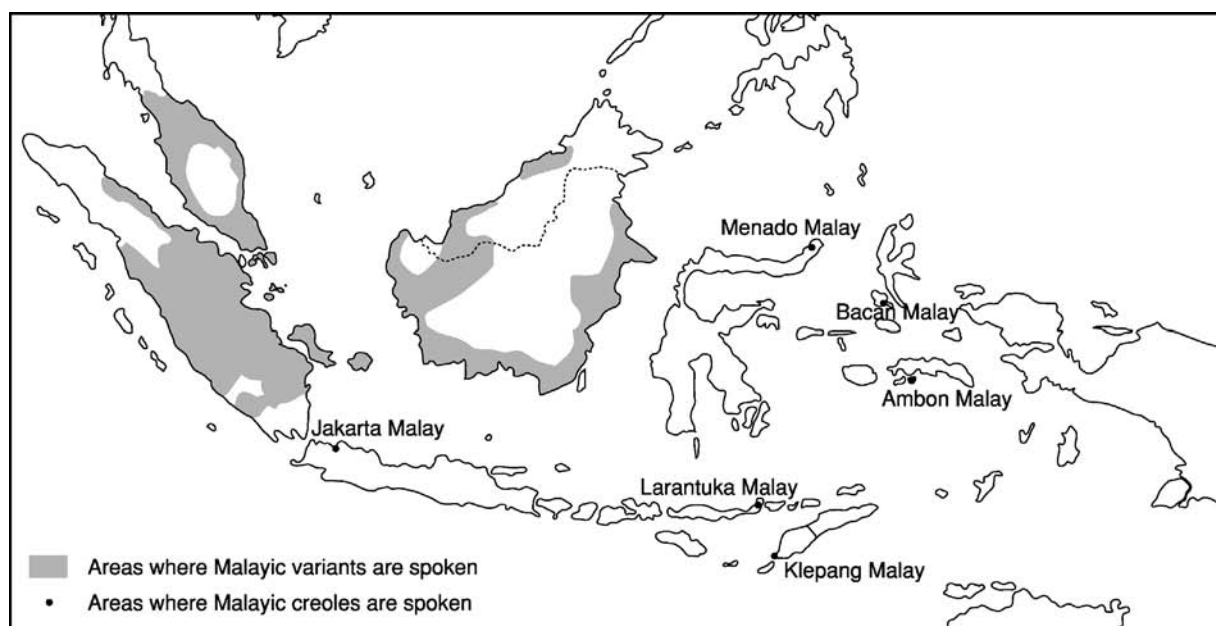


Figure 1 Map of the Malay-speaking area.

had to sail, was prone to become the means of communication of all those involved in commercial activities in the Indo-Malaysian archipelago. With the development of the spice trade, this language was carried all the way to the Moluccas and to the many other harbor towns of this archipelago. When the Portuguese arrived in the early 16th century, simplified forms of Malay had already spread east and developed into creoles replacing local languages (e.g., Kupang Malay, Ambon Malay, Larantuka Malay). On the Malay Peninsula and on the adjacent southern islands, Malay developed literary varieties at the various royal courts. The most prestigious one was the literary classical Malay of the Riau-Johore kingdom, which had its roots in the literary tradition of the earlier sultanate of Malacca (Sneddon, 2003; Prentice, 1978).

The existence of two standard varieties of Malay, namely Malaysian (called 'Bahasa Melayu' in Malaysia) and Indonesian ('Bahasa Indonesia'), is mainly the result of an agreement reached between the British and the Dutch, who in 1824 drew new boundaries of their colonial territories. The mainland part of the Malay-speaking area became part of the British realm, and Sumatra together with the offshore islands became part of the Dutch realm. The treaty divided the former Riau-Johore Sultanate into two separate entities, with Johore belonging to the British and the Riau archipelago belonging to the Dutch. Because of this political demarcation, the influential Riau-Johore variant of Malay was now spoken in two distinct territories, which were to become Malaysia and Indonesia. Since this prestigious Riau-Johore court language played a major role in the formation of the standard languages of both countries, Malaysian and Indonesian remained closely related and are dialects of one and the same language. The differences between the two are most obvious in the vocabulary. The phonological, morphological, and syntactic differences are few and not very significant. There are a considerable number of cases in which Malaysian borrowed an English word and Indonesian a Dutch word, e.g., *tayar* vs. *ban* 'tire' or *fius* vs. *sekering* 'fuse.' Other variations occur when one of the two national variants has borrowed a European word, while the other one is a retention or an innovation, e.g., Malaysian *dulang* 'tray' (retention) vs. Indonesian *baki* 'tray' (from Dutch *bakje*) or Malaysian *panggung wayang* 'cinema' (innovation) vs. Indonesian *bioskop* (from Dutch *bioscoop*). There are cases when both Malaysian and Indonesian share the same word but with minor phonetic variation, e.g., Malaysian *kerusi*, Indonesian *kursi* 'chair' (from Arabic *kursi*). In some instances the Malay word underwent different semantic changes, e.g.,

Malaysian *pusing* 'turn, revolve' has the meaning 'dizzy' in Indonesian. Furthermore, Malaysian has borrowed more from Arabic than Indonesian, while Indonesian has undergone considerable Javanese and Jakarta Malay influence.

In Indonesia, the establishment of Malay as the national language was not disputed; its choice was not regarded as favoring any one ethnic group, since ethnic Malays constituted no more than 10% of Indonesia's population. Furthermore, various forms of Malay had long been established throughout the Indonesian archipelago. In Malaysia, the situation was different. When Malaysia became independent in 1957, Malay became the national language and one of the official languages (the other is English). Malay became the only language of education. Since Malay was more or less the exclusive property of the Malays, who made up about 50% of the population, the Chinese and Indian population of Malaysia felt at a disadvantage. A change of the language name from Bahasa Melayu to Bahasa Malaysia was one of the compromises made to comfort the non-Malay population. Later, however, the name Bahasa Melayu was reintroduced. In Malaysia, English still plays an important role and today competes with Malay as the language of instruction; in 1993, English became the language of instruction in universities. The Malaysian government argued that this was done in the interest of science and technology (Sneddon, 2003). Since 2003, secondary schools have taught mathematics and sciences in English. The introduction of English as language of education was based on the government's observation that the knowledge of English among pupils and students had deteriorated dramatically. Many Malays are worried that juxtaposing Malay and English against each other will result in a new linguistic scenario and marginalize the original national language policy.

In 1984 Malay also became the national language of Brunei Darussalam in northeast Borneo and is also called Bahasa Melayu. In this country it is the sole official language. The standard language is lexically much closer to Malaysian. In addition to Bahasa Melayu the state of Brunei also has another Malay variant (Brunei). This variant constitutes the main lingua franca in the coastal regions of Brunei

Table 1 Vowel phonemes in standard Malay

| | Front | Central | Back |
|------|----------|----------|----------|
| High | <i>i</i> | | <i>u</i> |
| Mid | <i>e</i> | <i>ə</i> | <i>o</i> |
| Low | | <i>a</i> | |

Diphthongo: -ay, -aw.

Table 2 Consonant phonemes in standard Malay

| | Labial | Dental | Alveolar | Palatal | Velar | Glottal |
|-----------------|----------|----------|----------|----------|----------|----------|
| Voiceless stops | <i>p</i> | <i>t</i> | | <i>c</i> | <i>k</i> | |
| Voiced stops | <i>b</i> | | <i>d</i> | <i>j</i> | <i>g</i> | |
| Nasals | <i>m</i> | | <i>n</i> | <i>ɲ</i> | <i>ŋ</i> | |
| Fricatives | | | <i>s</i> | | | <i>h</i> |
| Liquids | | | <i>l</i> | | <i>r</i> | |
| Semivowels | <i>w</i> | | | <i>y</i> | | |

(Nothofer, 1991). Brunei has an official bilingual education policy that preserves the status of Malay but recognizes the importance of English by making it the medium of instruction from the upper primary school onward in almost all subjects.

Malay is also the national language of Singapore and one of its four official languages, along with English, Mandarin Chinese, and Tamil. Malay is a minority language, spoken by not more than 15% of the population. In southern Thailand, more than a million speakers use a Malay variant, Pattani Malay.

Cooperation between Malaysia and Indonesia resulted in the spelling reform of 1972, which removed the differences in the spelling of consonants, e.g., former Malaysian *ch* and Indonesian *tj* are now spelled *c*; former Malaysian and Indonesian *dj* are now spelled *j*. The cultural pact between the countries was intensified in 1972 with the establishment of a council known as the Language Council for Indonesia and Malaysia (MBIM). Its main tasks are to create a common scientific terminology and cooperate closely on matters pertaining to language. In 1986, Brunei Darussalam officially joined as a member of the Council, which took the new name MABBIM.

Malay Phonology

The description of the Malay phonology shown here is that of Standard Malay (SM), as defined by Adelaar (1992: 3). The vowel phonemes of SM are shown in Table 1, and consonant phonemes are shown in Table 2. The consonant /r/ is realized as a velar or uvular fricative and elided word finally by speakers of the traditional Malay areas. It is an apical flap or trill outside these areas and in official Indonesian (Adelaar, 1992: 8).

Malay Morphology

Malay prefixes include: *bər-* 'stative, habitual'; *məN-* 'active, agent focus'; *dɪ-* 'passive, patient focus'; *məmpər-/dipər-* 'causative'; *tər-* 'accidental state, involuntary, agentless, sudden'; and *pəN-*, *pər-*, *pə-* 'actor of the performance, instrument with which the action is performed, someone having a quality as a

characteristic.' The common Malay suffixes are: *-an* 'collectivity, similarity, object of an action, place where the action is performed, instrument with which the action is performed' *-kan* 'causative, benefactive'; and *-i* 'locative, repetitive, exhaustive.' Malay circumfixes include: *bər-* *-an* 'diffuse action, plurality of subject'; *kə-* *-an* (verbal) 'unintentional action or state, potential action'; *kə-* *-an* (nominal) 'nouns referring to a quality, abstract nouns, collectivity'; *pəN-* *-an*, *pər-* *-an* 'abstract nouns, place where the action is performed, goal or result of action.'

Bibliography

- Adelaar K A (1992). *Pacific Linguistics C-119: Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Australian National University.
- Adelaar K A (1995). 'Borneo as a cross-roads for comparative Austronesian linguistics.' In Bellwood P, Fox J J & Tryon D (eds.) *The Austronesians: historical and comparative perspectives*. Canberra: Research School of Pacific and Asian Studies, Department of Anthropology, Australian National University. 75–95.
- Adelaar K A (1996). 'Malay: the national language of Malaysia.' In Wurm S A, Mühlhäusler P & Tryon D T (eds.) *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*. Berlin: Mouton de Gruyter. 673–693.
- Collins J T (1998). *Malay, world language: a short history*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Grimes C E (1996). 'Indonesian: the official language of a multilingual nation.' In Wurm S A, Mühlhäusler P & Tryon D J (eds.) *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*. Berlin: Mouton de Gruyter. 719–727.
- Martin P W, Ozog C & Poedjosoedarmo G (eds.) (1996). *Language use and language change in Brunei Darussalam*. Athens: Ohio University Center for International Studies.
- Moeliono A & Grimes C E (1995). 'Indonesian (Malay).' In Tryon D T (ed.) *Comparative Austronesian dictionary: an introduction to Austronesian studies 1*. Berlin: Mouton de Gruyter. 443–457.
- Nothofer B (1991). 'The languages of Brunei Darussalam.' In Steinhauer H (ed.) *Papers in Pacific linguistics, A-81*. Canberra: Australian National University. 151–176.

- Nothofer B (1997). *Dialek Melayu Bangka*. Bangi: Penerbit Universiti Kebangsaan Malaysia.
- Omar A H (1978). 'The use of the Malaysian national language in a multilingual society.' In Udin S (ed.) *Spectrum: essays presented to Sutan Takdir Alisjahbana on his seventieth birthday*. Jakarta: Dian Rakyat.
- Prentice D J (1978). 'The best chosen language.' *Hemisphere* 22(3), 18–23; 22(4), 28–33.
- Prentice D J (1990). 'Malay (Indonesian and Malaysian).' In Comrie B (ed.) *The major languages of East and South-east Asia*, 2nd edn. London: Routledge. 185–207.
- Sneddon J N (1996). *Indonesian: a comprehensive grammar*. Sydney: Allen & Unwin.
- Sneddon J (2003). *The Indonesian language: its history and role in modern society*. Sydney: University of New South Wales.
- Steinhauer H (1980). 'On the history of Indonesian.' In Barentsen A A, Groen B M & Sprenger R (eds.) *Studies in Slavic and general linguistics*, vol. 1. Amsterdam: Rodopi. 349–375.
- Teeuw A (1959). 'The history of the Malay language.' *Bijdragen tot de Taal-, Land-en Volkenkunde* 115, 138–156.
- Vikør L (1983). 'Language policy and language planning in Indonesia and Malaysia.' In Svensson T & Sørensen P (eds.) *Indonesia and Malaysia. Scandinavian studies in contemporary society*. London: Curzon. 47–74.

Malayalam

B Gopinathan Nair, St Xavier's College,
Thiruvananthapuram, Kerala State, India

© 2006 Elsevier Ltd. All rights reserved.

Language and Speakers

Malayalam, a major literary language of South India with long traditions of literature and scripts, is the main language of the state of Kerala and of the islands of Lakshadweep, which are 200–400 km off the southwest coast of India. Malayalis have migrated to different parts of India and overseas, especially to Malaysia, Singapore, the United States, Canada, the United Kingdom, and Australia. The number of Malayalam speakers in India is 31.83 million. In Kerala, 96% of the total population is composed of the religious majority (comprising Hindus, 58.1%) and the religious minorities (Muslims, 21.3%; Christians, 20.6%); these groups mostly speak Malayalam. Linguistic minorities comprise 5.2% of the population. Kerala has the highest literacy rate in India (90.6% of the population). The number of dailies and periodicals in Malayalam in 2000 was 1505 (according to the *Manorama year book* in 2004).

Etymology and Variant Names

Malayāḷam is a combination of *mala* 'mountain' with any of the following terms: *aḷam* 'the place,' denoting 'the mountain country'; *ālam* 'depth,' representing 'the land that lies between the mountain and the deep ocean'; or *āl* 'man,' meaning 'mountain dweller.' The last term may convey the original meaning of *Malayāḷam*, denoting both 'the people,' depicted by word forms such as *malayāḷar*, *malayāḷi*,

and *malanāṭṭukāran*, and the region or country, as in the term *malanāṭu*. Early variants include *malayāḷma*, *malayāyma*, and *mlayāṇma*. *Malayāḷam* may be a later variant. *Lilatilakam*, a famous 14th-century work on the grammar and language of Malayalam, mentions only *kēraḷabbhāṣa* to denote the language.

Development of Literature

Malayalam flourished in Kerala amidst continuous contact and convergence with Sanskrit, Prakrit, and Pali; profusely borrowing lexical items from these languages in addition to incorporating loans from Arabic, Persian, Urdu, Syriac, Portuguese, Dutch, Hindi, and English. The early development of Malayalam was considerably influenced by Sanskrit, the language of scholarship, and Tamil, the language of administration; eventually, Malayalam evolved in written documents and literature. The Brahmin contact made profound impact in adapting several Indo-Aryan features into Malayalam. Malayalam has a recorded literary history of over eight centuries; the earliest document, the Vaḷappaḷḷi inscription of Rajasekhara, dates to the 9th century. The early literature developed through three different traditions: (1) the Tamil tradition of *pāṭṭu*, the classical songs depicted in the first literary work, *Ramacaritam*, (2) the Sanskrit tradition of *maṇipravāla*, a literary innovation portraying a harmonious blend of *bhāṣa* and *Samskr̥ita* (i.e., the native language and Sanskrit – for instance, *Vaiśikatantram*), and (3) the native tradition of producing folk songs and ballads predominantly concerning indigenous elements. *Bhasakautiliyam* is the earliest prose written in simple language. All three traditions belong to the

12th century. Modern Malayalam literature is rich in fiction, poetry, prose, drama, short stories, biographies, and literary criticism.

Writing System

The early Malayalam writing system had evolved from Vatteluttu, traceable to the Pan-Indian Brahmi script; this system continued for a long period, eventually adding symbols from Grantha script to represent Indo-Aryan loans. The writing is based on the concept of the *akṣara* 'graphic syllable,' wherein the graphic elements have to be read as units, although the individual vowels and consonants are easily recognizable. The script reformation implemented in the 1970s made a reduction of the less frequent conjunct consonants and combinations of the vowel *u* with different consonants, to make a simpler writing scheme. The orthography is largely phonemic, with separate script for each phoneme (with a few exceptions). The dental and alveolar single nasals (*n* and *ṇ*) are depicted by the same script, as are their long counterparts. The direction of writing for several scripts is clockwise. In a few cases, the direction is clockwise plus anticlockwise or vice-versa within a single letter. Malayalam scripts bear simple to complex allographic representations. Geminated consonants and heteroelemental consonant clusters are marked differently by writing the consonants side by side, or one above the other. Additionally, there are other combinations of consonants that seldom follow regular patterns in graphemic depiction. The six consonants *m*, *n*, *ṇ*, *r*, *l*, and *ḷ* in word-final positions have separate symbols for writing.

Grammatical Tradition

Malayalam grammatical tradition commenced with the 14th-century *Lilatilakam*. The European contributions in the early 18th century were of great importance, especially Hermann Gundert's *Malayāḷabhāṣavyākaraṇam* (1851, 1868). The 19th century saw the publication of grammatical treatises by a few native scholars, viz., George Mathan (*Malayālmayūṭe vyākaraṇam*), Kovunni Nedungadi (*Kēraḷakoumudi*; 1878), and some others, but the most widely used work was that of A. R. RajaRaja Varma, the *Kēraḷa pāṇinīyam* (1896). This was followed by L. V. Ramaswamy Iyer's profound contribution to various aspects of Malayalam linguistics in 1925. The past four decades have witnessed the production of considerable work based on modern linguistic theories and descriptive techniques applied both to various written texts belonging to different centuries,

ranging from *Ramacaritam* to the 16th-century *Adhyatma Ramayanam*, and to regional, caste, communal, and tribal dialects, with the ultimate goal of preparing a historical grammar for Malayalam that is still a desideratum (much of this work can be found in the Ph.D. dissertations by scholars in the Department of Linguistics, University of Kerala).

Dialect Variation

Malayalam dialect variations are discernible with respect to phonetic, phonological, grammatical, semantic, and lexical levels and in intonation patterns along the parameters of caste, community, region, social stratum, education, occupation, style, and register. The speech forms of Travancore, Cochin, South and North Malabar, and the Lakshadweep islands show considerable differences. Among the 48 tribal languages in the hilly tracts of Kerala, many are dialects of Malayalam and a few belong to one or the other of South Dravidian languages.

The first systematic dialect survey of Malayalam was based on a single speech community, the Ezhava/Tiyas groups living throughout Kerala; the survey was completed in 1968, demarcating 12 major dialect areas (Subramoniam, 1974). This was followed by the Nair and Harijan dialect surveys. About 600 dialect maps were prepared concerning the Ezhava/Tiyas and Nair castes, along with frequency charts of the variants, showing differences with regard to 300 diagnostic lexical items. Copies of these maps are preserved in the collections of the Department of Linguistics, University of Kerala, and the International School of Dravidian Linguistics, Thiruvananthapuram, Kerala. Among several dialect variations, the occurrence of *y* in place of *l* is commonplace, as in *palam* > *payam* 'fruit,' but the occurrence of *t* (*ketakku* 'east') is a rare feature found in the northern part of Kasargod. Initial *v/b* alternation, as in *v/barin* 'come' and *v/bāppa* 'father,' is a distinct feature of Muslim speech throughout Kerala and Lakshadweep, but in Cannanore district this change is found in the speech of other castes, demonstrating the overlap of caste, communal, and regional traits. Word-final *n/m* alternation is a feature of the Muslim dialect of Ernad and Lakshadweep, as in *nēran/nēram* 'time.' Present tense markers *-anRa* and *-uṇṇa*, as in *kottanRa* 'chops' and *iṭuṇṇa* 'places,' is a peculiarity of the PaRaya speech of Kasargod; *-aṇṇa*, as in *bayaṇṇa* 'comes,' is found in the Muslim dialect of Lakshadweep.

The literary dialect is almost uniform. The language that is used in newspapers, in mass media, and in formal situations, which is largely understood by the majority of the people irrespective of caste,

community, and region, is considered to be the standard variety. A standard colloquial is slowly evolving.

Genetic Affiliation

Malayalam shows affinity to Tamil, Kota, Toda, Irula, Badaga, Kodagu, Kannada, and Tulu, all of which belong to the South Dravidian branch of the Dravidian family. However, the affinity with Tamil is greater, since Malayalam emerged from Proto-Tamil–Malayalam; divergence occurred over a period of four or five centuries, from the 8th century onward, and distinct languages, separate from Tamil, were established.

Three distinctive features of Proto-Tamil–Malayalam include (1) *k* > *c*- before front vowels, whether followed by a retroflex or not, (2) **e*, **o* > *i*, *u* before a derivative suffix beginning with *a*, and (3) the presence of the accusative suffix *-ai*. The features that distinguish Malayalam from Tamil are (1) progressive assimilation of nasal + stop > nasal + nasal except in retroflexes and labials, (2) loss of person–number–gender in finite verbs, (3) negative periphrastic construction with *illa*, and (4) prohibitive construction with infinitive + *arutu*.

Characteristic Features

Phonology

Five vowels with length contrast, *i*, *e*, *a*, *o*, and *u*, occur in the literary and spoken dialects. Two diphthongs, *ai* and *au*, occur in the literary language. An onglide *y* and *v* occur word initially in front and back vowels, respectively, in pronunciation. Vowels occur in all positions except for short *o*, word finally. The *u* is pronounced as a high back rounded vowel when it occurs initially, medially without length in the first syllable, and with length finally, whereas it is pronounced as a lower high back unrounded vowel (*samvurutookaaram*) medially except in the first syllable and without length word finally.

Single voiceless stops (except in clusters with homorganic nasals) are pronounced with voicing (with or without slight fricativization) when occurring intervocalically; voiceless stops preceded by homorganic nasals are pronounced with slight voicing. Generally, aspirated plosives lose aspiration in pronunciation. Only six consonants, *m*, *n*, *ɳ*, *r*, *l*, and *ʃ*, can occur word finally; *n* occurs medially either with length or in clusters. All other consonants can occur word initially and medially. Voiceless stops occur medially with length and in clusters, except with homorganic nasals. For consonants, length contrast occurs only medially; *r*, *s*, *ś*, *ṣ*, *h*, and *ʈ* do not geminate.

Morphophonemics

The sandhi rules (a systematic blend of words; Sanskrit *sandhi* ‘to join’) fall under two categories, internal and external, the former operating within a word and the latter operating between words; the rules may operate in either category or in both categories (sandhis). For example, for a vowel *V*, the rule *V1 + V2(V2) + V2(V2)* operates only in close juncture: *āyi + illa* > *āyilla* ‘did not become.’ The following examples show other sandhi blends:

| | | | | | |
|-------|---|--------|---|--------------|---------------------------|
| i | | | i | | |
| e | + | v | > | e | + y + v (internal |
| a | | vv | | a | vv and external) |
| kutṭi | + | uṭe | > | kutṭiyuṭe | ‘child of’ |
| nalla | + | āḷu | > | nallayāḷu | ‘good person’ |
| tala | + | alla | > | talayalla | ‘head not’ |
| u | + | v(v) | > | u + v + v(v) | (internal and external) |
| rāmu | + | um | > | rāmuvum | ‘Ramuv also’ |
| kuru | + | āyi | > | kuruvāyi | ‘seed become’ |
| l/ṇ | + | n | > | ṇṇ | (external) |
| toḷ | + | nūRu | > | toṇṇūRu | ‘ninety’ |
| kaṇ | + | nīru | > | kaṇṇīru | ‘tears’ |
| m/n/ṇ | + | STOP | > | | (homorganic nasal + stop) |
| cem | + | tāmara | > | centāmara | ‘lotus’ |
| pin | + | tuṇa | > | pintuṇa | ‘support’ |
| peṇ | + | kutṭi | > | peṇkutṭi | ‘girl’ |

Morphology

Noun stems fall under three categories, viz., personal pronouns (first person, inclusive and exclusive), second person, and reflexives. Demonstrative base + gender number marker constitutes third-person pronouns, as in *av-an* ‘he,’ *avaḷ* ‘she,’ and *av-ar* ‘they.’

Numerals consist of adjectival and case bases. Number markers are *-n* (gender singular), *-m* (gender plural), *ṇān* ‘I,’ *nammal* ‘we (inclusive),’ *-tu* (non-gender-neutral singular), and *atu* ‘that.’ Examples of gender markers are *-n*, *-ān*, and *-an* (masculine) and *-ḷ*, *-atti*, and *-āṭṭi* (feminine). Plural suffixes are *-aḷ* and *-kaḷ*.

Nominative case has no marker; accusative uses *-e* and *-a*, dative uses *-u* and *-kku*, and instrumental uses *-aal* (in literary Malayalam; in dialects, postposition *kontu* ‘with’ is used). Sociative case uses *-ōṭu* and locative case uses *-il*.

Verbs do not distinguish person–number–gender. Both finite and nonfinite verbal forms consist of a verb stem followed by verbal suffixes, which take (or can take) tense markers. A few verbs do not take tense but can take negative markers (*illa* ‘no,’ *alla* ‘not,’ and *arutu* ‘do not’). Verbs fall into two groups, intransitive and transitive. Some of the former are transitivized morphologically in three ways:

(1) by suffixing the markers *-tt-* and *-kk-* to the intransitive verb stem (*iru* 'to sit,' *iru-tt-* 'to make to sit'; *oṭi* 'to break,' *oṭi-kk-* 'to make to break'), (2) by geminating the stem-final stops (*aaṭ-* 'to become,' *ākk-* 'to make to become'; *aaṭ-* 'to swing,' *aaṭṭ-* 'to make to swing'; *kēṛ-* 'to climb,' *kēRR-* 'to make to climb'), and (3) stem finally (nasal + nasal > homorganic stop + homorganic stop (*uRaññ-* 'to sleep,' *uRakk-* 'to make to sleep')). Two causative markers, *-i-* and *-ppi-*, can occur simultaneously within a verb, as in *paRay-i-ccu* 'caused to say,' *paRay-i-ppi-ccu* 'to cause to say.'

Three-way distinctions in tense occur, i.e., present, future, and past. Examples are *-unnu* (present tense) and *-um* (future tense), as in *var-unnu* 'comes' and *var-um* 'will come.' All vowel-ending stems take link morph *-kk-* before present and future markers, as in *paṭi-kk-unnu* 'learns' and *paṭi-kk-um* 'will learn.' The verb stem *vēṇ-* is peculiar in that it takes the future tense marker *-am* (*vēṇ-am* 'will need') but does not take either the present or the past tense suffix. There are several past tense markers: the vowel ending *-i* (*pāṭi* 'sang') and nasals *-nn-*, *-ññ-*, *-ṇ-*, and *-nt-* (*iru-nnu* 'sat,' *kara-ññu* 'wept,' *tā-ṇu* 'drowned,' *no-ntu* 'pained,' and *ve -ntu* 'boiled'); only these last two verbs take the past tense *-nt-*.

Stops are *-t-*, *-ṭ-*, *-R-*, and *-c-* (*eṭu-ttu* 'took,' *kaṇ-ṭu* 'saw,' *pe-RRu* 'delivered,' and *aṭi-ccu* 'beat'). Negative suffixes are *-ātt-* before the relative participle marker *-a* (*var-ātt-a* 'that which did not come'), *-āt-* before the verbal participle marker *-e* (*var-āt -e* 'having not come'), and *-a*, which freely varies with *-ā* (*vē-ṇa(a)* 'not needed'); *-ān* denotes the purposive infinitive (*paRay-ān* 'saying').

The vowel-ending stems can be used as imperatives (*paRa* '(you) tell,' *nōkku* '(you) look'), but are less polite in speech than *-ū* is (the more polite forms are *paRay-ū* and *nōkk-ū*). The optative marker is *-aṭṭe*, as in *var-aṭṭe* 'let (him) come.'

Syntax

Three major types of sentences, simple, complex, and compound, can be discerned. A simple sentence consists of the subject noun and predicate verb, as in *raaman varunnu* 'raman comes.' A nominal sentence in which both the subject and the predicate are nouns is seen in *atu maram* 'that is tree.' The finite verb *aanu* is optional. Malayalam word order is not rigid. Subject-object-verb is the usual order. A noun or noun phrase can be the subject in a sentence. A noun phrase (NP) is expandable by modifiers, the structure of which is \pm possessive \pm demonstrative \pm numeral Adj \pm Adj + NP, as in *enRe ā oru nalla*

peena 'my that one good pen' 'that good pen of mine.' A noun phrase can be expanded by a relative participle, nouns, case base, and clitics, as in *ceyta kāryam* 'thing done,' *kaykāryam* 'handling the affairs,' *tankaaryam* 'one's own affair,' *kāṭṭukōli* 'jungle fowl,' and *piRRe divas am* 'next day.'

Nouns/noun phrases can form the direct or indirect object. If the direct object is an animate noun, the accusative case suffix *-e* is added; if the direct object is an inanimate noun, the case suffix is dropped. The indirect object takes the dative case, as in *avaḷ sitaykku* (indirect object) *oru pūccaye* (direct object) *kot'uṭṭu* 'she gave a cat to Sita.' Verbs/verb phrases are expanded by verbal participles, auxiliary verbs, or adverbial clitics, as in *talayil cumatu veccukoṭuttu* 'placed a bundle on the head' and *patukke pooyi* 'slowly gone.'

Interrogative sentences can be formed by adding the interrogative clitic, which would yield 'yes' or 'no' types of answers (*avaḷ sitayānō* 'is she Sita?'). This can also denote doubt. After the defective verbs *illa* and *alla*, the interrogative particle *-ee* is added (*illee*, *allē*, 'is it not'). Interrogative words such as *aaru* 'who,' *eetu* 'which,' and *enttu* 'what' can be added to form sentences, as in *āru paRaññu* 'who told,' *ētu kāryam* 'which subject,' and *enttu vēṇam* 'what (do you) want.' Negative sentences are formed either by negativizing the verb phrase by using morphological negative markers, or by negation of the sentence or verb phrase by using defective verbs such as *illa* and *alla* (*pōyi* 'went,' *pōyilla* 'did not go,' *itu pēna āṇu* 'this is pen,' *itu pēna alla* 'this is not pen,' *āhāram uṇṭu* '(there) is food,' *āhāramilla* '(there) is no food').

Bibliography

- Andrewskutty A P (1971). *Malayalam, an intensive course*. Trivandrum: Dravidian Linguistic Association.
- Asher R E & Kumari T C (1996). *Malayalam*. London and New York: Routledge.
- Gopinathan Nair B (1971). 'Caste dialects of Malayalam.' In *Proceedings of the First All India Conference of Dravidian Linguists*. Trivandrum: Dravidian Linguistics Association. 409–414.
- Panikkar G K (1973). *Description of the Ernad dialect of Malayalam*. Trivandrum: Dravidian Linguistics Association.
- Somasekharan Nair P (1979). *Cochin dialect of Malayalam*. Trivandrum: Dravidian Linguistics Association.
- Subramoniam V I (1974). *Dialect survey of Malayalam*. *Ezhava/Tiya*. Trivandrum: University of Kerala.

Malayo–Polynesian Languages

M Ross, The Australian National University,
Canberra, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The term ‘Malayo–Polynesian’ today denotes the largest of the ten putative primary subgroups of the Austronesian language family (Blust, 1999). Malayo–Polynesian (MP) embraces perhaps 1100 languages, while the other nine groups consist only of the surviving fourteen Formosan languages of Taiwan (see **Formosan Languages**). Until Father Wilhelm Schmidt invented the term ‘Austronesian’ in 1899, however, ‘Malayo–Polynesian’ denoted the whole Austronesian language family. Its German equivalent, *malayisch-polynesisch*, was first used in print by Franz Bopp in 1841 (it is often wrongly attributed to Wilhelm von Humboldt, but is not found in his writings). ‘Malayo–Polynesian’ was in use with this meaning in English by the 1870s, but who first used it to refer to the language family is unclear (Ross, 1996). It continued to be used as a synonym for Austronesian until the 1970s, and is occasionally still used in this sense today.

In 1977 Robert Blust showed that the primary division of Austronesian was into several subgroups of languages spoken in Taiwan and a single subgroup which he labeled ‘Malayo–Polynesian’ and which includes all the Austronesian languages spoken outside Taiwan: the Austronesian languages of the Philippines, Southeast Asia, Madagascar, the Indo-Malaysian archipelago, New Guinea, Island Melanesia, Micronesia, and Polynesia. This is the sense of ‘Malayo–Polynesian’ in the remainder of this article. Some scholars prefer the term ‘Extra-Formosan’ in its place. Clearly the potential for overlap between this and a discussion of Austronesian languages is great, and the reader is referred to **Austronesian Languages** for further information.

The Integrity of the Malayo–Polynesian Subgroup

How do we know that all Austronesian languages outside Taiwan belong to a single subgroup? To determine a family tree we first compare the languages of the family and reconstruct the protolanguage from which they are descended, in this case Proto Austronesian (PAn). Then we identify subgroups of languages whose members share a set of innovations relative to PAn. We infer that the innovations are

shared because they have been inherited from a single interstage language. This is far more probable than the alternative assumption – that the innovations have occurred independently in each language that reflects them.

All Austronesian languages outside Taiwan reflect certain phonological innovations relative to PAn, and we infer that they occurred in a single interstage language which Blust named Proto Malayo–Polynesian (PMP). These innovations are enumerated here (from Blust, 1990) with minimal discussion and examples.

- A. PAn *t and *C merged as PMP *t.
- B. PAn *L and *n merged (with some unexplained exceptions) as PMP *n.
- C. PAn *S became a glottal spirant of some kind, but did not merge with *h.

Innovation A is illustrated below, where PAn *t and PAn *C (which remain separate in the Formosan language Rukai) are merged in MP languages, exemplified by Itbayat, a language of the Batanes islands between Taiwan and Luzon:

PAn *tuLa ‘freshwater eel’ (Rukai *tola*) > PMP

*tuna (Itbayat *tuna*)

PAn *pitu ‘seven’ (Rukai *pito*) > PMP *pitu (Itbayat

pitu)

PAn *Caliŋa ‘ear’ (Rukai *tsaliŋa*) > PMP *taliŋa

(Itbayat *taliña*)

PAn *maCa ‘eye’ (Rukai *matsa*) > PMP *mata

(Itbayat *mata*)

In innovation B, PAn *L and *n merged as PMP *n:

PAn *qaLup ‘hunt’ (Rukai *alopo*) > PMP *qanup

(Itbayat *anup*)

PAn *wanan ‘right (hand)’ (Rukai *vanan*) > PMP

*wanan (Itbayat *wanan*)

Innovation C is reflected in

PAn *duSa ‘two’ (Rukai *dosa*) > PMP *duha

(Itbayat *duha*).

A major set of innovations in pronouns involved a ‘politeness shift’ (Blust, 1977). Just as the English plural pronoun *you*, used as a polite form of address, eventually displaced singular *thou*, so PMP underwent a set of changes in pronouns which were also related to politeness (for details see Ross, 2002: 51). No MP language reflects forms that predate the shift.

PMP added to the verbal system the prefixes *paN-(distributive), *paR-(durative, reciprocal) and *paka-(aptative, potential) (Ross, 2002: 49–50). These are widely reflected in the languages of the Philippines and the western part of the Indo–Malaysian

archipelago, and are preserved in fossilized form in many languages elsewhere in the MP subgroup.

History and Subgrouping

How did it come about that all Austronesian languages outside Taiwan belong to a single subgroup while perhaps nine coordinate groups are represented in Taiwan itself? The obvious answer is that PAN was spoken in Taiwan and diversified into a group of languages there. Speakers of one of these languages left Taiwan, presumably for the northern Philippines. Their language underwent the innovations noted above, becoming the language we call PMP.

Archaeological dating suggests that the culture that spoke PAN flourished in Taiwan around 3000 B.C. and that the migration to the Batanes islands or Luzon which led to the genesis of PMP occurred around 2000 B.C. The descendants of PMP speakers evidently spread, mostly south and then eastward, at an astonishing speed, colonizing the Philippines, the Indo-Malaysian archipelago, parts of coastal New Guinea, and the Bismarck Archipelago in the northwest of Island Melanesia within about 500 years, by 1500 B.C. This history is reflected in the tree diagram of the Austronesian family (see Figure 1). This shows some 20 to 25 groups of western MP languages, spoken in the Philippines and the western part of the Indo-Malaysian archipelago, with outliers on Hainan, in the Vietnamese highlands, on the islands along the western coast of Thailand and Myanmar, and on Madagascar (see Figure 2). The migration of MP speakers to Madagascar was a much later event (Adelaar, 1991). Adelaar (2004) provides a listing of western MP groups which reflects current understanding. Although there are frequent references in the literature to 'Western Malayo-Polynesian,' there was

never a 'Proto Western MP,' as western MP languages as a whole share no innovations. The similarities among western MP languages, such as they are, reflect shared retentions from PMP. The fact that the tree shows so many coordinate branches reflects the rapidity with which MP speakers occupied the region. They were agricultural people – rice growers – and probably encountered little significant opposition from the small populations of hunter-gatherers who had previously occupied these territories.

There is reasonable evidence in the form of shared innovations that all MP languages in the regions labeled on the map as Central Malayo-Polynesian (CMP), South Halmahera/West New Guinea (SHWNG) and Oceanic are descended from a single language, shown on Figure 1 as Proto Central/Eastern Malayo-Polynesian (PCEMP) (Blust, 1993). However, the set of innovations that defines this grouping is not nearly as substantial as the set defining MP (see above), and we must infer that the period for which PCEMP speakers remained an integrated speech community was short. The conventionally accepted family tree of Austronesian languages (originally proposed by Blust, 1977) gives PCEMP two daughters, 'Proto CMP' and 'Proto Eastern MP.' The status of both is doubtful. There is agreement among scholars today that PCEMP diversified, apparently rapidly, into a dialect network, and that the Eastern Malayo-Polynesian languages broke away from that network, probably as the dialects of the network were achieving the status of separate languages. There is no significant evidence, however, that there was ever a discrete Proto CMP (for more details, see Ross, 1995).

The existence of Proto Eastern MP is also questionable, and it is possible that it was simply a peripheral section of the Central/Eastern MP dialect network

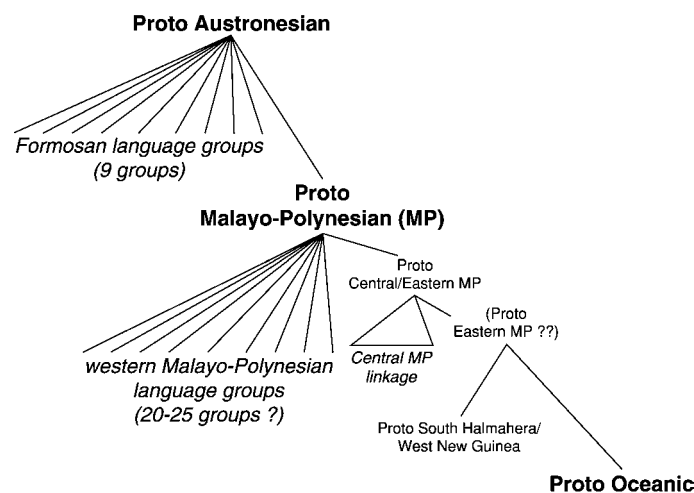


Figure 1 Austronesian family tree.

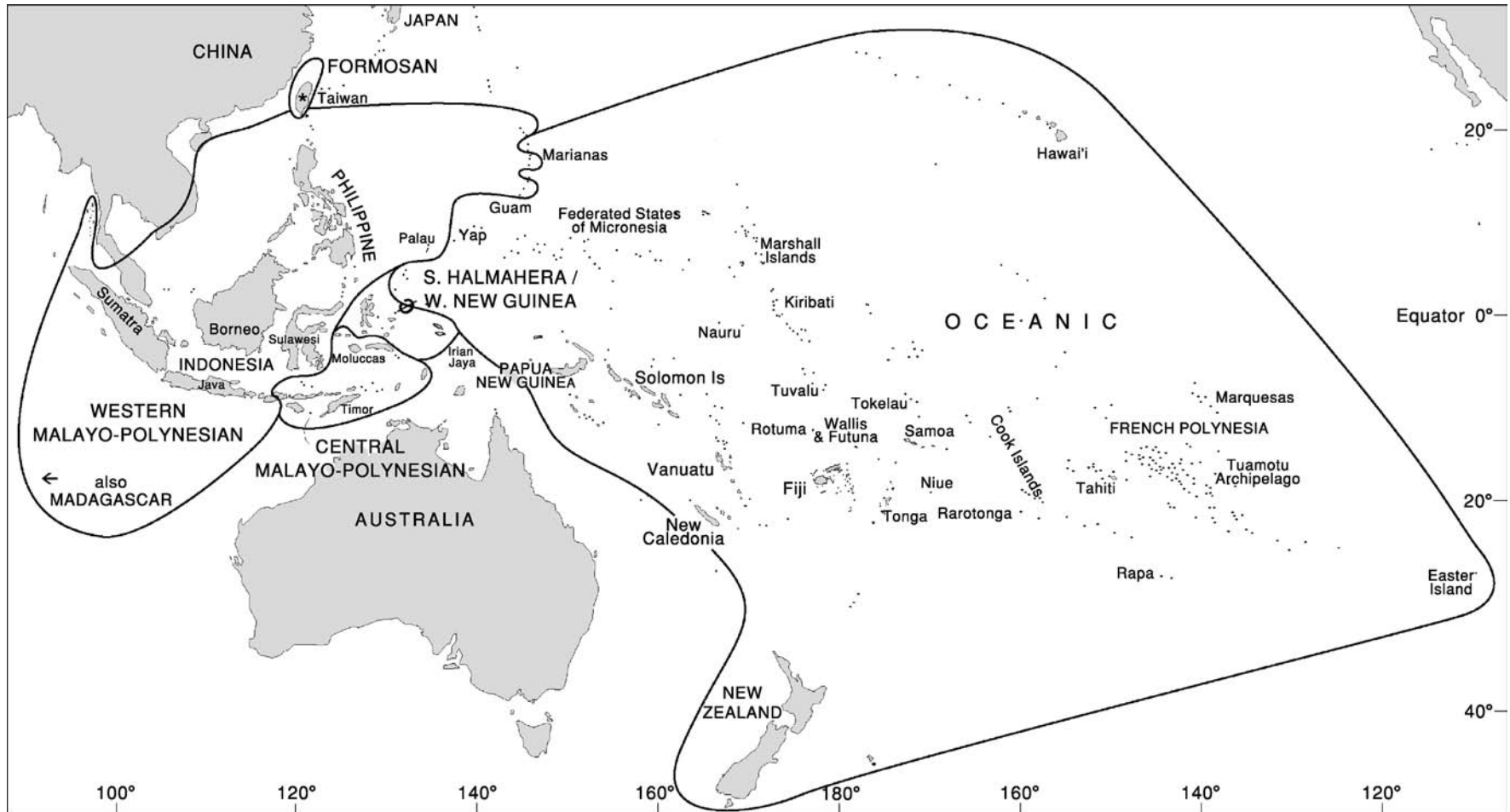


Figure 2 The Austronesian family and major Malayo-Polynesian language groups.

(Ross, 1995; Adelaar, 2004). However, there is much less doubt about Proto SHWNG, the ancestor of a small group of languages in the south of Halmahera and scattered around the Bird's Head Peninsula of New Guinea, and no doubt at all about Proto Oceanic.

Proto Oceanic was the ancestor of the MP languages of New Guinea other than those belonging to SHWNG (see map) and of all the MP languages of Island Melanesia, Polynesia, and Micronesia other than Chamorro (Guam) and Palauan (Belau). The Oceanic languages share a striking set of innovations, larger than the set for Proto MP itself. These innovations were first recognised by Dempwolff (1937) in Volume 2 of his pioneering work on Austronesian (see **Austronesian Languages**) and have undergone various modifications since as the result of further research (Lynch *et al.*, 2002: 63–67).

The history represented by the varying strengths of the nodes in the Austronesian family tree diagram (Figure 2) shows a period of relative stability during which Proto MP developed from the speech of those who emigrated from Taiwan, followed by 500 years of extraordinary settlement activity which culminated in the arrival of MP speakers in the Bismarck Archipelago. Here there seem to have been a few more centuries of relative stability, during which Proto Oceanic developed into a language that was at least phonologically and lexically rather different from its sisters around the Bird's Head.

Why the apparent halt in settlement activity? There were perhaps two reasons. First, New Guinea was already inhabited by Papuan speaking agriculturalists (see **Papuan Languages**) with much greater population densities than their hunter-gatherer neighbors to the west, and there was little space for the newcomers. Second, there probably was continued settlement activity during the development of Proto Oceanic, but no further than the Solomon Islands, to the east of which there is a substantial sea gap (Pawley, 1981).

There is agreement among many linguists and archaeologists (but not all) working in Island Melanesia that Proto Oceanic was the language of the Lapita Culture, a group that produced distinctive pottery and exploded eastwards into the Pacific from about 1300 B.C. Island Melanesia (the Bismarck Archipelago, Solomon Islands, Vanuatu, New Caledonia, and Fiji), Tonga, and Samoa were all settled within a few hundred years (Kirch, 1997). A linguistic puzzle in this story is that Proto Polynesian, the Oceanic language ancestral to Tongan, Samoan, and the 40 or so languages of scattered Polynesian communities is structurally rather different from other Oceanic languages, yet there is no

obvious hiatus in the archaeological record during which these differences might have developed. It is reasonably certain, however, that Proto Polynesian or an immediate ancestor developed in the northeastern islands of Fiji (Geraghty, 1983).

The Structures of Malayo-Polynesian Languages

MP languages show an extraordinary structural diversity. General accounts can be found in Lynch *et al.* (2002: 34–53) for Oceanic languages and Himmelmann (2004) for other MP languages. The languages of the Philippines and parts of northern Borneo, northern Sulawesi, and Madagascar largely retain the structure of PAn (see **Austronesian Languages**; **Formosan Languages**). The western MP languages of Vietnam, Hainan, and the Thailand/Myanmar islands show the influence of Mon-Khmer languages. In the western MP languages of Malaysia and western Indonesia we find a complex set of developments in which the PAn voice system is much reduced but applicatives take over much of its functional load (Ross, 2002). The CMP, SHWNG, and Oceanic languages (other than Polynesian) have a broad typological similarity, but with many variations. Most of these languages have lost all trace of voice and have subject-referencing verbal prefixes or proclitics. How this system arose from systems reflected in western MP languages is traced by Lynch *et al.* (2002: 57–63). Klammer (2002) describes CMP language structures, Ross (2004) those of non-Polynesian Oceanic languages.

Further References

The MP family is vast, and the serious enquirer will need to look beyond this article. Detailed maps of the locations of MP languages and their dialects are found in Wurm and Hattori (1981–1983), although it is not a reliable source for subgrouping. Adelaar and Himmelmann (2004) is the major reference for MP languages other than Oceanic, Lynch *et al.* (2002) for Oceanic languages. Both works also include a large collection of grammar sketches of a sample of languages. Tryon (1995) is an extensive comparative lexicon.

Associated with the historical study of MP languages, especially of Oceanic, is a solid body of work on culture history. Ross *et al.* (1998; 2003) are the first two of five volumes in which the terminologies used by Proto Oceanic speakers are reconstructed, following more piecemeal work on the lexicons of MP languages in general (Pawley and Ross, 1994) and work on MP culture history by scholars in various disciplines (Bellwood *et al.*, 1995). Pawley and

Ross (1993) is a short survey mostly of MP historical linguistics and cultural history.

Bibliography

- Adelaar K A (1991). 'New ideas on the early history of Malagasy.' In Steinhauer H (ed.) *Papers in Austronesian linguistics no. 1*. Canberra: Pacific Linguistics. 1–22.
- Adelaar K A (2004). 'The Austronesian languages of Asia and Madagascar: a historical perspective.' In Adelaar & Himmelmann (eds.) 1–42.
- Adelaar K A & Himmelmann N (eds.) (2004). *The Austronesian languages of Asia and Madagascar*. London: Routledge.
- Bellwood P S, Fox J J & Tryon D (eds.) (1995). *The Austronesians: historical and comparative perspectives*. Canberra: Department of Anthropology, Research School of Pacific and Asian Studies, the Australian National University.
- Blust R A (1977). 'The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report.' *University of Hawaii Working Papers in Linguistics* 9(2), 1–15.
- Blust R A (1993). 'Central and Central-Eastern Malayo-Polynesian.' *Oceanic Linguistics* 32, 241–293.
- Blust R A (1999). 'Subgrouping, circularity, and extinction: some issues in Austronesian comparative linguistics.' In Zeitoun E & Jen-kuei Li P (eds.) *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academia Sinica. 31–94.
- Dempwolff O (1937). *Vergleichende Lautlehre des Austronesischen Wortschatzes*. (Vol. 2). *Deduktive Anwendung des Urindonesischen auf Austronesische Einzelsprachen*. Berlin: Dietrich Reimer. *Beihefte zur Zeitschrift für Eingeborenen-Sprachen* 17.
- Geraghty P (1983). *The history of the Fijian languages*. Honolulu: University of Hawaii Press.
- Himmelmann N (2004). 'The Austronesian languages of Asia and Madagascar: typological characteristics.' In Adelaar & Himmelmann (eds.) 110–181.
- Kirch P V (1997). *The Lapita peoples: ancestors of the Oceanic world*. Oxford: Blackwell.
- Klamer M A F (2002). 'Typical features of Austronesian languages in Central/Eastern Indonesia.' *Oceanic Linguistics* 41, 363–383.
- Lynch J, Ross M & Crowley T (2002). *The Oceanic languages*. Richmond, Surrey: Curzon.
- Pawley A K (1981). 'Melanesian diversity and Polynesian homogeneity: a unified explanation for language.' In Hollyman J & Pawley A (eds.) *Studies in Pacific languages and cultures in honour of Bruce Biggs*. Auckland: Linguistic Society of New Zealand. 269–309.
- Pawley A & Ross M D (1993). 'Austronesian historical linguistics and culture history.' *Annual Review of Anthropology* 22, 425–459.
- Pawley A & Ross M D (eds.) (1994). *Austronesian terminologies: continuity and change*. Canberra: Pacific Linguistics.
- Ross M D (1995). 'Some current issues in Austronesian linguistics.' In Tryon (ed.) 1, 45–120.
- Ross M (2002). 'The history and transitivity of western Austronesian voice and voice-marking.' In Wouk F & Ross M (eds.) *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics. 17–62.
- Ross M (2004). 'The morphosyntactic typology of Oceanic languages.' *Language and Linguistics* 5, 491–541.
- Ross M (compiler) (1996). 'On the origin of the term "Malayo-Polynesian."' *Oceanic Linguistics* 35, 143–145.
- Ross M, Pawley A & Osmond M (eds.) (1998). *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society* (Vol. 1). *Material culture*. Canberra: Pacific Linguistics.
- Ross M, Pawley A & Osmond M (eds.) (2003). *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society* (Vol. 2). *The physical environment*. Canberra: Pacific Linguistics.
- Thurgood G (1999). *From ancient Cham to modern dialects: two thousand years of language contact and change*. Honolulu: University of Hawai'i Press.
- Tryon D T (ed.) (1995). *Comparative Austronesian dictionary*. Berlin: Mouton de Gruyter.
- Wurm S A & Hattori S (eds.) (1981–1983). *Language atlas of the Pacific area*. Part 1: *New Guinea area, Oceania, Australia*. Part 2: *Japan area, Taiwan (Formosa), Philippines, mainland and insular Southeast Asia*. Canberra: Australian Academy of the Humanities.

Maltese

J Cremona

© 2006 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 5, pp. 2356–2357, © 1994, Elsevier Ltd.

Maltese is the national language of the Republic of Malta and one of its two official languages, the other being English. It is spoken by virtually all the 345 418

(1985) inhabitants (plus ca. 80 000 Maltese immigrants in Australia). Until the 1930s, its status was low, with the prestige languages being Italian and English. The first text written in Maltese, a poem, is ca. 1460 AD, but although texts appear sporadically thereafter, Maltese only began to be written systematically from about the end of the eighteenth century.

Maltese, a language of Arabic origin, shares many of the features that distinguish the modern Arabic

vernaculars from literary Arabic. Maltese also displays those features that distinguish Maghrebine dialects from the rest; for example, the loss of gender distinctions in second-person-singular pronouns and verbs and the leveling of first-person markers in the imperfect to give {n ... ø} for the singular and {n ... u} for the plural. However, Maltese differs from most 'core' vernaculars of Arabic by having (a) adopted the Roman alphabet; (b) a phonemic system without emphatics, with fewer back consonants but more vowels; (c) virtually the whole of the vocabulary pertaining to intellectual, technical, and scientific pursuits taken from Sicilian, Italian, and English; (d) a number of conservative lexical features (e.g., *ra* 'he saw'); and (e) grammatical innovations of Romance origin (e.g., passives with *kien* 'he was' or *gie* 'he came' as auxiliaries). These features reflect the fact that Maltese has

not had Classical Arabic as an acrolect for some seven centuries. This and the fact that it is now the national language of an independent state have given Maltese the status of a distinct language.

Sample: Il-gimgha l-oħra sibt ruħi, għall-ewwel
/il 'dʒima l 'ohra sɪpt 'ru:hi all 'ewwel
Last week I found myself, for the first

darba f'hajti, f' 'lecture theatre' ta' l-Università.
'darba f 'hajti f 'lektʃər 'ti:ætcər ta l universi'ta/
time in my life, in a lecture theatre of the University.

Bibliography

Isserlin B S J (1990). 'The Maltese language.' In Bosworth C E, Van Donzel E, Lewis B & Pellat C (eds.) *The encyclopaedia of Islam*. Leiden: E. J. Brill. 295–298.

Malukan Languages

M Florey, Monash University, Clayton, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

Some 128 languages are spoken in the geopolitical region of the Malukan islands in eastern Indonesia (see Figure 1). The majority of the 111 Austronesian languages of Maluku are subgrouped within the Central Malayo-Polynesian branch of Central-Eastern Malayo-Polynesian (CEMP) (Blust, 1978). A number of the Austronesian languages of north Maluku

are subgrouped in the South Halmahera-West New Guinea branch of CEMP (Collins and Voorhoeve, 1983). Current information indicates that there are also 17 non-Austronesian languages spoken in Maluku (Grimes, 2000). Sixteen West Papuan phylum languages are found in the northernmost parts of Maluku, on Morotai, Ternate, Tidore, Halmahera, and nearby smaller islands. Oirata is a Trans-New Guinea phylum language of southern Kisar island, located near the north-eastern tip of East Timor.

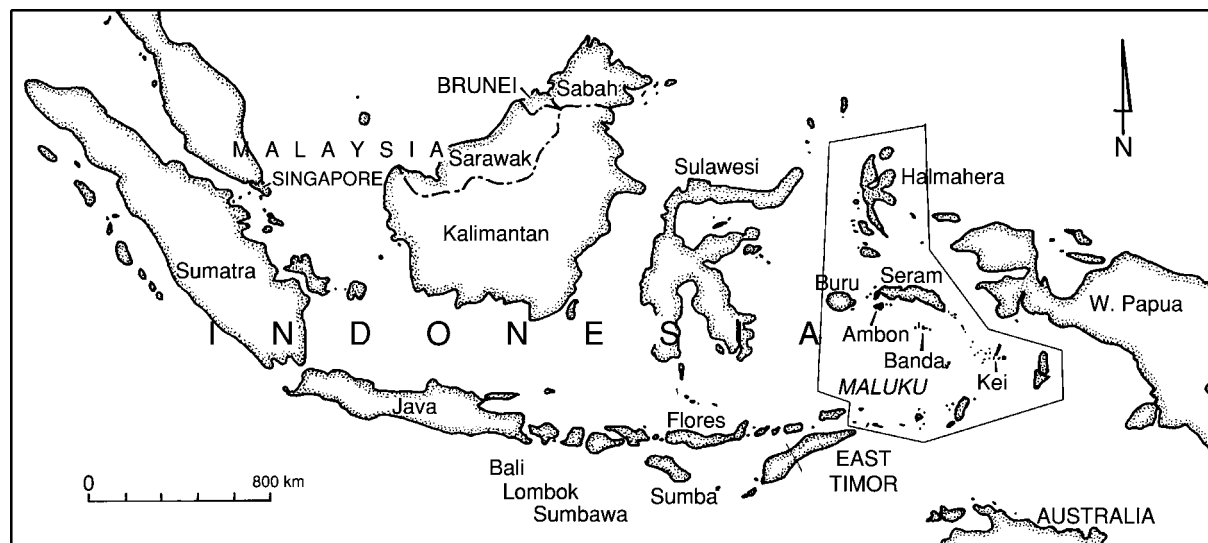


Figure 1 Map of Indonesia showing the location of the Malukan islands.

Linguistically, Maluku is characterized by high linguistic diversity, serious endangerment, and little detailed documentation. Speaker populations have historically been much smaller than in ethnolinguistic communities in the western Austronesian region. Larger languages include Kei with 86 000 speakers and Buru with perhaps 43 000 speakers (Grimes, 1995). These numbers, however, give an overly optimistic picture of linguistic vitality. The highest documented degree of language endangerment in Indonesia is located in Maluku. A recent survey indicated that 10 languages are close to extinction and a further nine languages are seriously endangered (Florey, 2005). Centuries of contact with nonindigenous peoples through colonization and intensive trade for spices, and conversion to non-indigenous religions have all played a role in language endangerment, which is particularly severe in the central Malukan islands of Seram and Ambon. Language contact has resulted in the wide use of a number of Malay creoles throughout Maluku of which Ambonese Malay is the best known (Minde, 1997).

Tryon (1994: 12) suggested that Maluku is possibly the least known Austronesian area and many languages – both Austronesian and non-Austronesian – remain undescribed. The richest descriptions to date include those of Alune (Florey, 1998, 2001) and Buru (Grimes, 1991, 1995) in central Maluku, Taba (Bowden, 2001) and Tidore (Staden, 2000) in north Maluku, and Leti (Engelenhoven, 1995) in south Maluku. These descriptions provide some insights into oral genres, including origin tales, historical narratives, folktales, riddles, and incantations. Parallelism (paired correspondences at the semantic and syntactic levels) is a feature of incantations and some narrative genres. Among the special registers which have been documented are those which were associated with avoidance relationships, hunting, fishing, healing, and headhunting. In some communities, ritual language still accompanies ceremonies held to mark the passing of life stages, and ritual practices associated with agriculture, renewing inter-village alliances, and the building of ritual houses.

Comparative analysis indicates that, in central Malukan languages, the preferred word order within clauses is SVO. Actor arguments in Alune may occur as a full noun phrase, a pronoun, or a proclitic, and actor NPs and pronouns are optionally crossreferenced with a proclitic on the verb. Undergoer arguments may occur as a full noun phrase, a pronoun, or an enclitic.

Au beta-ʔu-ru
1SG opp.sex.sibling-1SG.POSS.INALIEN-PL
esi-tneu behe a-ʔeri-ʔe sarei

3PL-ask CMP 2SG-work-APP what
'my younger siblings they asked me: "What did you work at?"' (Alune AK: 45)

This pattern of crossreferencing is not always apparent today as the rapid language change which accompanies language endangerment is typically characterized by extensive variation both within and between speech communities.

A morphologically marked alienable–inalienable contrast has been described for a number of the languages of Central Maluku. Synchronically, this contrast is not found across all languages. In those languages in which the contrast is marked, inalienable possession denotes all items which are culturally considered to be intrinsically a part of oneself – the things which we as humans are born with, and certain physical and emotional states. Alienable possession denotes the things which we might acquire through our lives: certain relationships and objects or possessions. Inalienable possession is marked with enclitics and alienable possession with proclitics, as demonstrated in the following Haruku examples:

Au oi kura au ama-ʔu
1SG go with 1SG father-1SG.POSS.INALIEN
kura au ina-ʔu
and 1SG mother-1SG.POSS.INALIEN
'I went with my father and my mother'

Esi-kana esi-lapu-na
3PL-fetch 3PL.POSS.ALIEN-shirt-NM.PL
lalu ani reu
then.MAL 1PL return.home
'they fetched their shirts then we went home'

Bibliography

- Blust R A (1978). 'Eastern Malay-Polynesian: a subgrouping argument.' In Wurm S A & Carrington L (eds.) *Proceedings of the Second International Conference on Austronesian Linguistics*. Canberra: Pacific Linguistics.
- Bowden J (2001). *Taba: description of a South Halmahera language*. Canberra: Pacific Linguistics.
- Collins J T (1983). *The historical relationships of the languages of Central Maluku, Indonesia*. Canberra: Pacific Linguistics.
- Collins J T & Voorhoeve C L (1983). 'Moluccas (Maluku).' In Wurm S A & Hattori S (eds.) *Language atlas of the Pacific area 2: Japan area, Taiwan (Formosa), Philippines, mainland and insular south-east Asia*. Canberra: Australian Academy of the Humanities in collaboration with the Japan Academy.
- Engelenhoven A van (1995). *A description of the Leti language (as spoken in Tutukei)*. Ridderkerk, Netherlands: Offsetdrukkerij Ridderprint.
- Florey M (1998). 'Alune incantations: continuity or discontinuity in verbal art?' *Journal of Sociolinguistics* 2(2), 205–231.

- Florey M (2001). 'Verb and valence in Alune.' *La Trobe Papers In Linguistics* 11, 73–120.
- Florey M (2005). 'Language shift and endangerment.' In Adelaar A & Himmelmann N P (eds.) *The Austronesian languages of Asia and Madagascar*. London: Routledge. 43–64.
- Florey M & Kelly B F (2002). 'Spatial reference in Alune.' In Bennardo G (ed.) *Representing space in Oceania: culture in language and mind*. Canberra: Pacific Linguistics. 11–46.
- Florey M & Wolff X Y (1998). 'Incantations and herbal medicines: Alune ethnomedical knowledge in a context of change.' *Journal of Ethnobiology* 18, 39–67.
- Grimes B F (ed.) (2000). *Ethnologue: languages of the world* (14th edn.). Dallas: SIL International.
- Grimes C E (1991). 'The Buru language of eastern Indonesia.' Ph.D. dissertation, Australian National University.
- Grimes C E (1995). 'Buru (Masarete).' In Tryon D T (ed.) *Comparative Austronesian dictionary: an introduction to Austronesian studies*. Berlin: Mouton de Gruyter. 623–636.
- Minde D van (1997). *Malayu Ambong: phonology, morphology, syntax*. Leiden: Research School CNWS.
- Staden M van (2000). *Tidore: a linguistic description of a language of the North Moluccas*. Delft: Systeem Drukkers.
- Steinhauer H (forthcoming). 'Endangered languages in south-east Asia.' In Collins J T & Steinhauer H (eds.) *Endangered languages and literatures in south-east Asia*. Leiden: KITLV Press.
- Tryon D T (1994). 'The Austronesian languages.' In Tryon D T (ed.) *Comparative Austronesian dictionary: an introduction to Austronesian studies 1*. Berlin: Mouton de Gruyter. 5–44.

Mambila

B Connell, York University, Toronto, ON, Canada

© 2006 Elsevier Ltd. All rights reserved.

Mambila is a Bantoid language situated in the Nigeria-Cameroon borderland. Mambila is a diverse language with approximately 20 different dialects. Among its interesting characteristics is its system of four level tones, and in one lect the presence of two fricative vowels that appear to be reflexes of the so-called super-close vowels of Proto-Bantu. Several Mambila lects are endangered, with some on the verge of extinction.

Classification

Mambila has been recognized since the early 1960s as a Bantoid language. A subgrouping within Bantoid now known as Mambiloid, which includes a number of other languages in the region, was proposed a decade later, although the precise relationship between Mambiloid and the rest of Bantoid remains a matter of debate.

Mambila is the most diverse of the Mambiloid languages. It is spoken on both sides of the Nigeria-Cameroon border, on the Mambila Plateau in Nigeria, and on the western edges of the Adamawa Plateau and the Tikar Plain in Cameroon. The great majority of Mambila speakers – an estimated 90 000 of 100 000 total speakers – are in Nigeria. Mambila comprises some 20 dialects, which are divided into two clusters, referred to by their rough geographical orientation as East and West Mambila. Within each

cluster there is limited mutual intelligibility among dialects, reflecting a dialect continuum; between the two clusters, mutual intelligibility does not exist, although speakers do recognize the relatedness of their languages to other languages in the region. Strictly on the basis of linguistic criteria, one might be inclined to refer to many of these dialects as distinct languages. For this reason, the neutral term 'lect' is used in referring to individual varieties of Mambila.

The main characteristic distinguishing the two dialect clusters is a difference in morpheme structure: In East Mambila a disyllabic root structure, CVCV(C), predominates, which corresponds to a monosyllabic CVC structure in West Mambila. A number of sound correspondences also serve to distinguish the two groupings, e.g., initial /f/ and /h/ in East Mambila correspond to /p/ and /t/, respectively, in West Mambila.

Little descriptive work has been done on Mambila. The two lects that have received the greatest attention are Tungba, spoken in Nigeria, and Ba, in Cameroon. Both are West Mambila lects. The following paragraphs present a short summary of Mambila structural characteristics.

Phonology

Consonants

Across Mambila lects there is little difference among consonant systems; what differences do exist are mostly related to the historical developments described above. The Ba system, /p, b, t, d, k, g, \widehat{kp} , \widehat{gb} , m, n, η , $\widehat{\eta m}$, mb, mv, nd, nd₃, ηg , $\widehat{\eta mgb}$, f, v, s, h, l, j,

w/, is fairly typical, both in its inventory and in the fact that /p, kp, gb, and ɣmgb/ in those lects, where they do occur, are infrequent. Distribution of consonants within the word is skewed in Mambila, with all consonants occurring in the initial position, but typically only /p, t, k, m, n, ɲ, l/ being found word finally.

Vowels

There is greater variation in the vowel systems of Mambila lects than for the consonants. The vowel system found in Ba, /i, e, a, ə, u, o, ɔ/ is the smallest; Tungba has only a slightly larger system, /i, e, ɛ, a, u, o, ɔ, ɑ/, but its phonetic realization is divergent, with allophonic front rounded vowels. Len, another West Mambila lect, is even more divergent, particularly with the presence of two fricative vowels, /ɬi, vuu/. These vowels appear to be the result of sub- or adstratal influence from the neighboring Grassfields languages, which may ultimately reflect the super-close vowels of Proto-Bantu.

Tone

Mambila is a register tone language; in West Mambila, it features four level tones that function both lexically and grammatically, and tones combine on single syllables to form a number of surface contours. Pitch realization in the Ba lect has been the subject of a number of experimental phonetic studies (Connell, 1999, 2000b, 2002). Tone in East Mambila lects has not been systematically investigated, although it is known that they have only three level tones.

Morphology

Mambila marks grammatical functions through affixation, typically suffixation. In most West Mambila dialects many of these functions are indicated only with a tonal morpheme; comparative evidence from both West and East Mambila reveals that a –CV melody is reconstructible in most cases. Despite the fact that Mambila is a Bantoid language, there is no system of nominal classification, and only traces remain of a former noun class system that reveals the heritage shared with Bantu. Pluralization is marked through means of a segmental suffix, –bV, except in Ba and other lects on the Tikar Plain, where a cognate prefix is used. There is evidence of an older means of marking plurals, and the presence of the common –bV is likely a recent development, perhaps through areal influences.

Syntax

Little can be said at this point concerning the syntax of Mambila. It has basic SVO word order, which

varies to indicate narrative, focus, and other pragmatic functions. As mentioned above, tone is used to indicate a number of grammatical functions, including negation, imperatives, and discontinuous verb phrases.

Language Vitality

Since it has more than 100 000 speakers, one might expect that Mambila will remain relatively stable for the foreseeable future. However, when its internal dialect variation is considered, an average of approximately 5000 speakers exists for each lect. Many of these are spoken by considerably fewer speakers; indeed, a few Mambila lects are on the verge of extinction, while one other has just recently become extinct. Given the potential contribution that these lects could make not only to our understanding of our common linguistic heritage but also to the history and prehistory of sub-Saharan Africa, documentation of these lects must be considered a priority.

Bibliography

- Blench R (1993). 'An outline classification of the Mambiloid languages.' *Journal of West African Languages*, 23(1), 105–118.
- Connell B (1999). 'Four tones and downtrend: a preliminary report on pitch realization in Mambila.' In Kotey Paul (ed.) *Trends in African linguistics 3: New dimensions in African linguistics and languages*. Trenton, N.J: Africa World Press. 75–88.
- Connell B (2000a). 'The integrity of Mambiloid.' In Wolff E & Gensler O (eds.) *Proceedings of the 2nd World Congress of African Linguistics, Leipzig*. Cologne: Rüdiger Köppe. 197–213.
- Connell B (2000b). 'The perception of lexical tone in Mambila.' *Language and Speech* 43, 163–182.
- Connell B (2002). 'Tone languages and the universality of intrinsic F0: evidence from Africa.' *Journal of Phonetics* 30, 101–129.
- Perrin M (1974). 'Direct and indirect speech in Mambila.' *Journal of Linguistics* 10, 27–37.
- Perrin M (1974). 'Mambila.' In Bendor-Samuel J (ed.) *Ten Nigerian tone systems*. Jos: Institute of Linguistics. 93–108.
- Perrin M (1994). 'Rheme and focus in Mambila.' In Levinsohn S (ed.) *Discourse features of ten languages of West-Central Africa*. Dallas: Summer Institute of Linguistics and University of Texas at Arlington. 231–241.
- Zeitlyn D (1993). 'Reconstructing kinship or the pragmatics of kin talk.' *Man* 28(2), 199–224.
- Zeitlyn D & Connell B (2003). 'Fractal history on an African frontier: Mambila – Njerep – Mandulu.' *Journal of African History* 44, 117–138.

Manambu

A Y Aikhenvald and P Y L Laki, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

Manambu belongs to the Ndu language family, and is spoken by about 2500 people in five villages: Avatip, Yawabak, Malu, Apa:n and Yuanab (Yambon) in East Sepik Province of Papua New Guinea. Between 200 and 400 speakers live in the towns of Port Moresby, Wewak, Lae, and Madang. Most Manambu speakers are proficient in Tok Pisin, the *lingua franca* of Papua New Guinea; many know English. In terms of number of speakers, the Ndu family is the largest in the Sepik area, comprising 32% of the Sepik basin dwellers (Roscoe, 1994). It consists of at least eight languages spoken by over 100 000 people along the course of the middle Sepik River and to the north of it. Other documented languages in the family are: Abelam or Ambulas (ca. 40 000; this number includes speakers of a variety of dialects under the names of Maprik, Wosera, West Wosera, and Hanga Kundi); Boikin (ca. 30 000); Iatmul (ca. 12 000); Sawos (ca. 9000); Yelogu (ca. 200); and Ngala (ca. 130). No genetic links between Ndu and other languages of the Sepik area have been proved. The origins, protohome, and the internal classification of the Ndu languages remains a matter for debate. Manambu's closest relatives are Iatmul and Ngala. The trade relationship and marriage exchange with the Iatmul contributed to a large amount of lexical diffusion between the two groups in close contact.

Manambu is synthetic, agglutinating with some fusion, mostly suffixing, and predominantly verb-final. The phonology of Manambu is complicated, with 21 consonants, nine vowels, and contrastive stress. Nouns distinguish eight cases (subject, definite object/locative; dative/aversive; allative/instrumental; comitative; terminative 'up to the point'; and two cases referring to 'means of transport'). Three numbers (singular, dual, and plural) and two genders (feminine and masculine in the singular) are expressed via agreement on demonstratives, interrogatives, in possessive constructions, on verbs and on two adjectives ('big' and 'small'). Singular and plural numbers are marked on kinship nouns, and on a few nouns from other semantic groups. The noun 'child' has a semisuppletive form for the dual number. Associative plural is marked on kinship nouns and personal names, as in *Tanina-b,r* 'Tanina and others.' Gender is distinguished in second and third person singular independent pronouns, and neutralized in the plural.

Nouns are assigned genders according to the sex of a human referent, and to shape and size of any other referent. That is, men are assigned to the masculine, and women to the feminine gender; a large house is masculine, and a small house feminine. By semantic extension, an unusually big or bossy woman can be treated as masculine, and a squat fattish man as feminine. Personal names are a distinct subclass of nouns, with special derivational suffixes not used anywhere else in the grammar.

Verbs have a plethora of grammatical categories, covering person, number, gender, tense, numerous aspects (e.g., completive, habitual, and repetitive) and modalities including irrealis, purposive, desiderative, and conditional. A verb in the declarative mood can cross-reference the person, number, and gender of the subject. Or, if a clause contains a constituent that is more topical than the subject, this constituent can also be cross-referenced alongside the subject. The imperative mood also marks person and number of the subject employing a different set of markers. The only fully productive prefix in the language is *a-*, the marker of second person imperative. Three suffixes expressing prohibition differ in their illocutionary force. Many of the verbal categories – including person and tense – are neutralized in negative clauses. Verb compounding is highly productive; up to three verbal roots can occur together, but the meaning of the combination is frequently unpredictable. Directionality (up, down, inside, outside) is marked both on verbs and on demonstratives. In addition, demonstratives encode six degrees of distance and visibility.

Similarly to other non-Austronesian languages of New Guinea, Manambu has extensive clause-chaining and a complex system of switch-reference, whereby a nonfinal clause is marked differently depending on whether its subject is the same, or differs, from that of the main clause. See Aikhenvald with Laki (forthcoming) for a full account of Manambu grammar, and also Aikhenvald (1998) and Allen and Hurd (1972). The relative complexity of Manambu could be partially accounted for by the substrata of languages spoken by members of neighboring tribes conquered by the Manambu as a result of inter-tribal warfare (Harrison, 1993).

Manambu culture places particular importance on ownership of personal names and various kinds of cultural knowledge. Ritualized debates among rival leaders and the clan groups they represent are, traditionally, the main political forum, and ownership of names is an oft-debated issue. A detailed study of Manambu ethnography is in Harrison (1990, 1993), which also contains a detailed analysis of the kinship

system and relationships (of Siouan type). Traditional genres include mourning songs *grakudi* and foiled marriage songs *namai* (Harrison, 1982; Takendu, 1977).

Manambu is an endangered language. All the Manambu are bilingual in Tok Pisin (and some also know English). Children in the villages prefer using the local *lingua franca*, Tok Pisin, in their day-to-day interaction. A literacy program in Manambu is currently being implemented at the local school at Avatip.

Bibliography

- Aikhenvald A Y (1998). 'Physical properties in a gender system: a study of Manambu.' *Language and Linguistics in Melanesia* 27, 175–187.
- Aikhenvald A Y with Laki P Y L (Forthcoming). *The Manambu language from East Sepik, Papua New Guinea*.
- Allen J D & Hurd P W (1972). 'Manambu phonemes.' *Te Reo* 15, 37–44.

Farnsworth R & Farnsworth M (1966). *Manambu grammar sketch 1: Morphology*. Ukarumpa: Summer Institute of Linguistics.

Harrison S (1982). *Laments for foiled marriages: love songs from a Sepik River village*. Port Moresby: Institute for Papua New Guinea Studies.

Harrison S (1990). *Stealing people's names: history and politics in a Sepik river cosmology*. Cambridge: Cambridge University Press.

Harrison S (1993). *The mask of war*. Manchester/New York: Manchester University Press.

Laycock D (1965.). *The Ndu language family (Sepik District, New Guinea)*. Canberra: Linguistic Circle of Canberra Publications.

Roscoe P (1994). 'Who are the Ndu? Ecology, migration, and linguistic and cultural change in the Sepik Basin.' In Strathern A J & Stürzenhofecker G (eds.) *Migration and transformations: regional perspectives on New Guinea*. Pittsburgh: University of Pittsburgh Press. 49–84.

Takendu D (1977). 'Avatip village, Ambunti sub-province, East Sepik.' *Oral History* 5(5), 2–53.

Mande Languages

D Dwyer, Michigan State University, East Lansing, MI, USA

© 2006 Elsevier Ltd. All rights reserved.

The Distribution of the Mande-Speaking People

Today, the Mande (or *Mandé* in French) language group consists of some 30 languages spoken in West Africa from Nigeria to Senegal by an estimated 10 million speakers. The term Mande and its variants (see Table 1) provide not only the basis for many of the names of the Northern Mande languages, but the various names accorded the language family as well. These variants are attributable to (1) minor vowel alternations (*e/i* and *a/e*), (2) the consonantal alternation (*nd/n/l*) found throughout the Mande-speaking area, and (3) the suffix *-ka(n)*, meaning 'language or dialect.'

Table 1 Variants of Mande

| | |
|---------------|--------------------|
| Mandi(ng) | Mandinka, Mandingo |
| Mande (Mandé) | Mandekan |
| Mende | |
| Mani | Manianka |
| Mane | |
| Mali | Malinke |
| Male | |

The map of the distribution of the Mande languages (Figure 1) shows that the heaviest concentration of Mande languages is in the republics of Guinea and Mali, and adjacent areas of Senegal, Sierra Leone, Liberia, and Ivory Coast. Furthermore, these western languages are contiguous and cover larger areas than those to the east, which appear as islands in a sea of Niger-Congo languages.

The Reconstructed History of the Mande

While scholars have not reached a total consensus on how the Mande evolved, evidence from the historical, archaeological, and linguistic record suggests the following six stages.

Phase 1: The Drying of the Desert

According to McIntosh and McIntosh (1984), the Mande originally lived in a much wetter Saharan area and practiced a herding–fishing–collecting economy. Lexicostatistical evidence (Dwyer, 1989) suggests that 4000 years ago the Mande people were undifferentiated linguistically.

Around 3000 B.P. (before present), in response to the increasing lack of rainfall, one branch of Proto-Mande (the earliest form of Mande) speakers migrated southward where wetter conditions would permit their herding–fishing–collecting way of life. The other

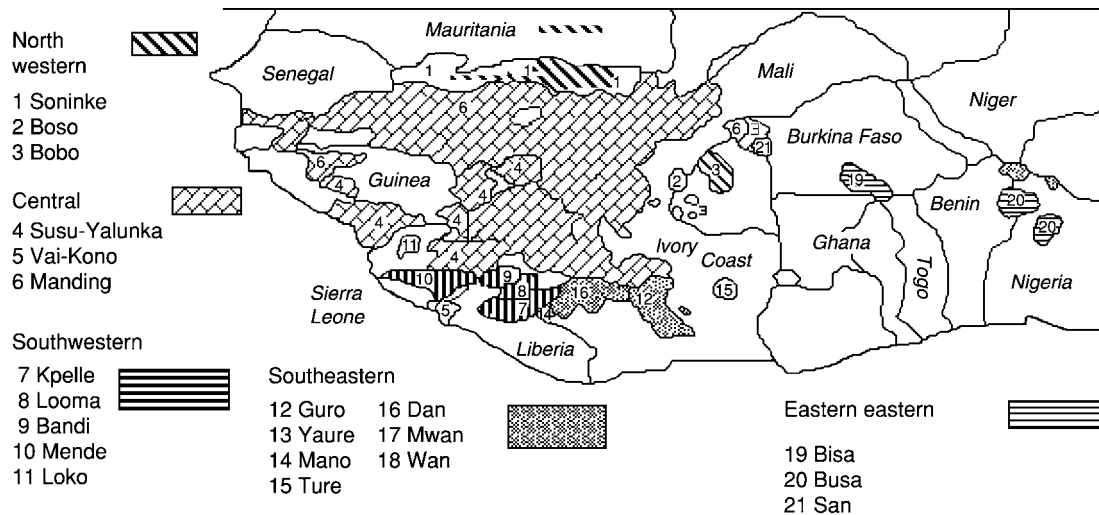


Figure 1 Map of the Mande languages.

branch, known as the western branch, responded to the increasing dryness by intensifying their cultivation of cereals.

Phase 2: The Development of Agriculture

At Jenno-Jene in the upper Niger delta, archaeologists have identified a site, continuously occupied from about 2250 B.P., that exhibits a second agricultural phase. This elaboration of agriculture may well have been responsible for the diversification and the westward expansion of the Central Mande speakers. This expansion may also have been responsible for pushing the pastoral Soninke further to the north.

Phase 3: The Rise of the Sudanic Kingdoms of Ghana and Mali

At the time of the sedentarization of the Western Mande, the people of this area were engaged in extensive trans-Saharan trade with North Africa. The stimulus for the trade was the alluvial gold found in deposits along the upper Niger River, which was exchanged for Mediterranean merchandise and salt. This trade gave rise to the Soninke-speaking empires of Ghana (700–1100) and the Manding-speaking empire of Mali (800–1550). The substantial area taken up by the Western Mande can be attributed to the expansion of this empire.

Phase 4: Rice and the Development of Forest Agriculture

While research in this area is still in progress, evidence suggests that a form of upland rice in the Guinea Highlands and iron tools permitted the Mande (and Atlantic) populations living along the rainforest-savannah border to enter the forest to practice swidden agriculture. The map shows a number of

Mande groups straddling this line including the Southwestern Mande, the Vai/Kono into present-day Sierra Leone and Liberia, and many of the Eastern Mande peoples into Liberia and Ivory Coast. Using oral traditions and genealogies, Person (1961) concludes that this movement into the rain forest took place in the 15th century. As these agricultural people moved into the sparsely populated rain forests, they increased the risk of malarial infections. In response to this situation, the percentages of the sickle cell trait (an adaptation to malaria) increased in these populations to the point where they are among the highest in the world (Livingstone, 1958).

Phase 5: The Arrival of the Europeans

Beginning with the Portuguese in 1455, contacts, trade, and finally settlement in this area increased, so that by 1500 permanent trading outposts and slaving operations were fully established. One effect of this development was the decreasing economic importance of the trans-Saharan trade and the decline of the Sudanic kingdoms.

The Linguistic Evidence

Following a technique developed by Ehret (1980), Table 2 shows the Proto-Mande terminology relating to economic activity (hunting, herding, and agriculture). After establishing the lexicostatistical dates, vocabulary common to all or most members of the branch are considered to be in existence at the time the branch was an undifferentiated language. Thus the terms for wine, mortar, and dog were common to the western branch but not the eastern branch, and are presumed to have been part of Western Mande before it separated into its constituent

groups. This linguistic evidence is consistent with that proposed for the early phases of Mande.

The Classification of Mande Languages

Current Classification

The internal classification of Mande (see Table 3) has undergone a series of revisions, the most recent and most accurate being that done by Kastenholtz (1996). For a full classification of the Mande languages go to the Ethnologue website.

Earlier Classifications

Mande was first recognized as a related group of languages by Sigmund Koelle, who used the

term Mandinga (Koelle, 1854). Shortly thereafter Heymann Steinthal (1867) introduced the term Mande (or *Mandé*). Maurice Delafosse offered the first subclassification of Mande in 1901, in which the major distinction was between Mande Tan (which is the northern group minus Susu and Yalunka) and Mande Fou, based on the words for 'ten.' Over time, the Tan/Fou categorization became increasingly suspect, but it was not until William Welmers (1958), using a lexicostatistical approach based on the Swadesh 100-word basic vocabulary list, rejected it and produced the first version of the currently accepted system. Welmers concluded that the word *tan* was a more recent innovation in Western Mande, not the fundamental split that Delafosse had assumed, and introduced the East–West division that remains today.

Mande as a Niger-Congo Language

Since the time of Koelle, four major hypotheses concerning the placement of Mande in Niger-Congo have been offered:

- Westermann (1927) included Mande in his West Sudanic, which was very similar to Greenberg's Niger-Congo (Table 4).
- In 1963, Joseph Greenberg, using a methodology based on the mass comparison of lexical items, accepted and refined Westermann's view. He renamed West Sudanic as Niger-Congo, the name it bears now, and included it as a branch of a larger grouping, Niger-Kordofanian. Of all the Niger-Congo languages, Greenberg considered Mande the least remote (Table 5).
- Consistent with common usage, Williamson (1977) replaced Greenberg's Niger-Kordofanian with the

Table 2 Linguistic evidence

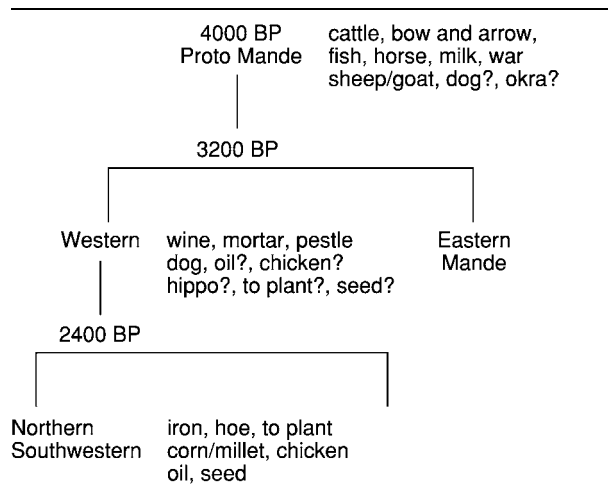


Table 3 Current classification

| Mande 4000 BP ^a | | | |
|--------------------------------|----------------------|-----------------------|------------------|
| West Mande 3200 BP | | Eastern Mande 2400 BP | |
| Central (southwestern) 3200 BP | Northwestern 2400 BP | Eastern-eastern | Eastern southern |
| South western 3000 BP | Jowulu | Bisa (Bissa) | Guro-Tura |
| Kpelle | Soninke-Bobo 2750 BP | Barka | Guro |
| Mende Looma (Loma) | Bobo | Lebir | Yaouré |
| Central 2100 BP | Soninke | Busa | Tura (Toura) |
| Susu-Yalunka | | Boka (Boko) | Dan |
| Manding-Jogo | | Bokabaru (Bokobaru) | Mano |
| | | Busa-Bisa (Busa) | Nwa-Ben |
| | | Tyenga (Kyenga) | Ben |
| | | Sam (Samo) | Gban (Gogu) |
| | | San (Samo) | Nwa (Wan) |
| | | Sane (Samo) | Mwan |

^aThe BP dates are from Dwyer (1989). Each date represents the estimated date at which the languages in the group separated, based on common percentages of basic vocabulary cognates. Thus Central Mande, for example, with its time depth of 2100 BP, is based on a common cognate percentage of 40%.

Table 4 Westermann*West Sudanic*

| | | | | | |
|---------------|-------|-----|--------------|-----|-------------------|
| West Atlantic | Mande | Gur | Togo Remnant | Kwa | Benue-Cross River |
|---------------|-------|-----|--------------|-----|-------------------|

Source: Westermann (1927).

Table 5 Greenberg

| |
|-------------------|
| Niger-Kordofanian |
| Niger-Congo |
| West Atlantic |
| Mande |
| Gur |
| Togo Remnant |
| Kwa |
| Benue-Cross River |
| Kordofanian |

Source: Greenberg (1963).

Table 6 Williamson

| |
|--------------------|
| Niger-Congo |
| Kordofanian |
| Mande |
| Atlantic-Congo |
| Atlantic |
| North |
| Bijago |
| South |
| Volta-Congo |
| Kru |
| Kwa |
| Benue-Congo |
| Dogon |
| Adamawa-Gur-Ubangi |
| Ijoid? |

Source: Williamson (1977).

term Niger-Congo. Williamson then placed Mande along with Atlantic Congo (the main body of Niger-Congo languages) and Kordofanian as the first three branches of Niger-Congo (Table 6).

- Also in 1977, Hans Mukarovsky proposed a substantial restructuring in which Mande and Benue-Congo were removed from the old Niger-Congo (renamed West Sahelian) and placed with Songhai (Songhay), previously not considered a Niger-Congo language, as branches of Sahelian (Table 7).

Although the Mukarovsky model is still seen as an interesting hypothesis, currently most scholars favor the Williamson proposal. Nevertheless, the progression from Westermann to Williamson to Mukarovsky does show an increasing awareness of Mande as a

remote branch of Niger-Congo. This development has raised questions about whether Mande is actually a Niger-Congo language. Part of this suspicion is due to the fact that Mande is also unique among the Niger-Congo languages because of its lack of evidence of a noun class system, found in other Niger-Congo languages, and its almost universal subject-object-verb word order.

This led Dwyer (1998) to compare the vocabulary of Mande and samples from other Niger-Congo branches. This study shows that Mande is a lexically coherent group. By lexically coherent, I mean that the best way to explain the vocabulary basic and other is to attribute a common ancestor (Proto-Mande) to these languages. The study also found that Niger-Congo (specifically Mande, Benue-Congo (including Bantu) and the western Nigritic core) is also lexically coherent. Finally, the study concluded that of these three language groups, Mande is lexically least related. These conclusions are fully consistent with the Williamson hypothesis but not that of Mukarovsky.

Linguistic Properties

Phonology

A tentative reconstruction of the Proto-Mande consonant system (Table 8) suggests a series of labial, alveolar, velar, and labiovelar voiced and voiceless stops. Because of the eccentric, but relatively consistent bimodal patterning of the voiceless stops, Dwyer (1994) tentatively suggested the possibility of a second series of fortis voiceless stops (t' , k' , kp'). Interestingly, this dual series of voiceless stops is analogous to that postulated for Upper Cross by Dimmendaal (1978) and Sterk (1979) and for Volta-Congo, Stewart (1976). In addition Mande appears to have only an (s/z) fricative contrast and labial, alveolar, and palatal nasals along with the liquid (l) and the glides (y and w).

Tone

Most Mande languages have two level lexical tones (high and low), along with a falling tone, analyzed as a sequence of high followed by low, and a rising tone. Bobo (Bobo Madaré), Mano, and Kpelle have three tones and one language, Sembla (Seeku), has four. Both Kpelle and Bobo (Bobo Madaré) (Dwyer, 1994)

Table 7 Mukarovsky

| <i>Sahelian</i> | | | |
|--|----------------------|-------------|------------------------------|
| <i>West Sahelian [Niger-Congo minus Mande and Benue-Congo]</i> | | | <i>[Mande-Benue-Songhai]</i> |
| <i>West Atlantic</i> | <i>West Nigritic</i> | <i>Kwa</i> | |
| Senegalian | Mel | Western Kwa | Mande |
| Mel | West Guinean | Eastern Kwa | Benue-Congo |
| West Guinean | Togo Remnant | | Songhai |
| | Gur | | |
| | Western Kwa | | |
| | Eastern Kwa | | |

Source: Mukarovsky (1976–1977).

Table 8 Mande consonants

| | <i>Labial</i> | <i>Dental</i> | <i>Palatal</i> | <i>Velar</i> | <i>Labiovelar</i> |
|--------------|---------------|---------------|----------------|--------------|-------------------|
| Stop | p/b | t/d | | k/g | kp/gb |
| Fricative | | s/z | | | |
| Nasal | m | n | ɲ | | |
| Liquid/Glide | | l | y | | w |

can be shown to have independently evolved a third tone through tone splitting. This suggests that they originally had a two-tone system.

Morphosyntax

One of the most striking facts about the Mande languages is the structural unity of the group and its distinctiveness from other Niger-Congo languages. Syntactically, the Mande languages have an SOV word order with oblique objects being marked as the objects of specialized postpositions. None of the Mande languages use serial verbs. Many Mande languages distinguish between alienable and inalienable possession.

Tense and aspect are generally marked through a combination of verb suffixes and postsubject formatives. Definite articles, demonstratives, and plurals tend to follow the noun or noun + attribute while possessive pronouns precede the noun.

Research in the area of comparative morphology and syntax is beginning to emerge. Creissels (1980) charted the distribution of four verbal particles in his Mandekan dialects with the conclusion that from these data no clear evolutionary sequence could be ascertained. Grégoire (1980) compared the rather unique properties of Mande relative clauses from all of its major branches: northern, southwestern, south-eastern, and Bobo. Dwyer (1985) has traced the evolution of the definite articles in Northwestern Mande.

Comparative reconstruction is a far more challenging task than lexicostatistical analysis, but promises

more interesting results, not only in the study of the development of the language, but also in the area of cultural history and in understanding the relationship between synchronic and diachronic rules.

Noun Classes

Typically, Niger-Congo languages have several semantically based noun classes (animate, inanimate, diminutive, augmentative, and abstract), usually marked by prefixes, both singular and plural. Languages of the Mande branch do not make use of this morphological device. One possible explanation is that the noun class system developed after Mande separated from Niger-Congo. Alternatively the system could have been part of Niger-Congo and subsequently lost in Mande. In the latter situation, one would expect some evidence in some of the Mande languages of remnants of such a noun class system. However, despite numerous attempts no evidence has turned up. For example, Dwyer (1990) examined Bobo (Bobo Madaré), which has a very complex system of plural formation requiring the positing of a number of noun classes in order to derive the correct form. However, these noun classes did not turn out to be related (semantically or morphologically) to the Niger-Congo noun classes.

A number of Mande languages have developed writing systems, including a Vai syllabary (Stewart, 1972) that has been in use continuously since the 1830s.

Resources

The Mande Studies Association (MANSA) website has several useful links to other resources in French (*Actualités de la recherche au Mali* and the *Bulletin d'Anthropologie et d'Histoire Africaines en Langue Française*) and English. Both the Summer Institute of Linguistics and Ethnologue contain descriptions of individual Mande languages and more detailed

maps. The site of the Union Mandingue has posted a history of the Manding-speaking peoples. Additional sources can be found by entering the individual language names given in this article in any search engine. The most thorough bibliography can be found in Kastenholz (1988).

Bibliography

- Bendor-Samuel J (ed.) (1989). *Niger-Congo*. Boston: University Press of America.
- Creissels D (1980). 'Variations dialectales dans les systèmes de marques prédicatives des parlers manding.' In Guarisma & Platiel (eds.), 139–155.
- Delafose M (1901). *Essai de manuel pratique de la langue Mandé ou Mandingue*. Paris: Leroux.
- Dimmendaal G (1978). *The consonants of Proto Upper-Cross and their implications for the classification of Upper Cross languages*. Leiden: Department of African Linguistics.
- Dwyer D (1989). 'Mande.' In Bendor-Samuel (ed.), 47–65.
- Dwyer D (1990). 'A second look at Bobo plurals.' Colloquium on African Linguistics, University of Leiden, September 5–7.
- Dwyer D (1994). 'Is there tone splitting in Bobo?' *Journal of African Languages and Linguistics* 15(1), 29–45.
- Dwyer D (1998). 'The place of Mande in Niger-Congo.' In Maddieson I & Hinnebusch T J (eds.) *Language history and linguistic description in Africa*. Trenton, NJ: Africa World Press.
- Ehret C (2002). *The civilizations of Africa: a history to 1800*. Charlottesville: University Press of Virginia.
- Grégoire C (1980). 'Les Structures relatives en mandé.' In Guarisma & Platiel (eds.), 139–155.
- Greenberg J (1966). *The languages of Africa*. Bloomington: Indiana University Press.
- Guarisma G & Platiel S A (eds.) (1980). *Dialectologie et comparatisme en Afrique Noire*. Paris: SELAF.
- Kastenholz R (1996). *Sprachgeschichte im West-Mande: Methoden und Rekonstruktionen*. Cologne: Rüdiger Köppe.
- Kastenholz R (1988). *Mande Languages and Linguistics*. Hamburg: Helmut Buske.
- Koelle S W ([1854] 1963). *Polyglotta Africana*. Graz: Akademische Druck-u. Verlagsanstalt.
- Livingstone F (1958). 'Anthropological implications of sickle cell gene distribution in West Africa.' *American Anthropologist* 60, 533–562.
- McIntosh R & McIntosh S (1984). 'Early iron age economy in the inland Niger delta.' In Clark D & Brandt S (eds.) *From hunters to farmers*. Berkeley: University of California Press.
- Mukarovsky H (1976–1977). *A study of Western Nigritic* (2 vols). Vienna: Institut für Ägyptologie und Afrikanistik.
- Sterk J (1979). 'The fortis/lenis contrast in upper cross consonants: a survey.' *Kiabará* special issue.
- Stewart J M (1976). *Towards Volta-Congo reconstruction*. Leiden: Leiden University Press.
- Stewart G (1972). 'The early Vai script as found in the book of Ndole.' Paper presented to the Conference on Manding Studies, School of Oriental and African Studies, University of London.
- Welmers W (1971). 'Mande Niger Congo.' *Current Trends in Linguistics* 7, 113–140.
- Westermann D (1927). *Die westlichen Sudansprachen und ihre beziehungen zum Bantu*. Berlin: Walter de Gruyter.
- Williamson K (1989). 'Niger-Congo Overview.' In Bendor-Samuel (ed.), 4–45.

Relevant Websites

www.ethnologue.com – Ethnologue.
www.manding-benelux.org – Union Mandingue.
www.sil.org – Summer Institute of Linguistics.
www.swt.edu – Mande Studies Association (MANSA).

Maori

W Bauer, Wellington, New Zealand

© 2006 Elsevier Ltd. All rights reserved.

Maori is the language of the Polynesian people who settled in New Zealand over 1000 years ago. It belongs to the Eastern Polynesian branch of the Malayo-Polynesian language family. Its current situation is typical of indigenous languages subjected to the effects of European colonization.

The early English missionaries Samuel Marsden and Thomas Kendall were instrumental in having a

writing system devised for Maori. This has largely served the language well, although it failed to distinguish between long and short vowels. The dictionary produced by three generations of the Williams family remains highly significant, and W. L. Williams made a substantial contribution to the grammatical description of Maori.

In the first half of the 20th century Maori children were taught English at the expense of Maori. Influential Maori leaders advocated the use of English in Maori homes, and speaking Maori at school was often punished. By the mid-20th century Maori was

rapidly dying, although small Maori-speaking communities remained in isolated rural areas. In 1951, Auckland University introduced Maori as an academic subject, which raised its status a little as did the grammatical descriptions produced by the linguist Bruce Briggs. However, the future looked bleak.

In the 1970s, a Maori political revival began (the 'Maori renaissance'). It was accompanied by serious endeavors to revitalize the language. *Kohanga reo* 'language nests' were established (preschool education centers providing education in Maori), and these were followed by Maori medium schools (*kura kaupapa Māori*), or immersion or bilingual units in mainstream schools. Small Maori radio stations were established with variable amounts of broadcasting in Maori. The Maori language became an official language of New Zealand. A Maori Language Commission was established to aid revitalization, and it has manufactured many vocabulary items from Maori elements to cater to modern needs. A Maori television channel went on the air in 2004.

Today the future of Maori is unclear. It might survive, testimony to the success of the revitalization process, but it is not yet secure. Most speakers have learned Maori as a second language, and many are 'semi-speakers.' Many teachers of Maori are not fully fluent, and the quality of the Maori taught in some Maori-medium classrooms is poor. Many children who leave *kohanga reo* speaking fluently do not use the language as teenagers. Most native speakers of Maori are over 70 years old, although there are some (particularly from the Tuhoe area, where Maori remained strong 20 years longer than elsewhere) still in the workforce. It is common to hear 'relexicalized English' – English structures where the content words and some of the grammatical words are replaced by Maori lexical items. Fluent Maori speakers often speak English to each other rather than Maori. While the latest surveys suggest that more people are speaking more Maori, there are very few who are fully conversant with the language. The last New Zealand census contained a question about use of Maori.

Such dialect differences as exist are tribally based. Most are lexical and phonological (with divergent realizations of phonemes rather than different phonological systems), and grammatical differences are not very significant.

Maori has a small phoneme inventory, with ten consonants (/p t k m ŋ r w f h/), and five vowels (/a e i o u/), each of which may be short or long. Orthographically, all use the obvious single roman letters except for <ng> (for /ŋ/) and <wh> (for /f/). The proper analysis of the long vowels is debatable: they may be analyzed as clusters of identical short vowels (reflected in 'double vowel' orthography), or

as separate phonemes (reflected in the macron orthography, which is now official): *Maaori* vs. *Māori*. Syllables are open, and there are no consonant clusters. All pairs of short vowels can occur in clusters, but some behave as one syllable and some as two. There are also longer vowel clusters. The rhythm is based on the mora (of form (C)V, where V is short), but stress operates with a bigger unit (C)V(V). Word stress is predictable, with syllables ranked according to the nature of the vowel. All content words contain at least two morae.

Maori has virtually no inflectional morphology and very little derivational morphology, although the allomorphs of the passive suffix have raised significant linguistic interest.

The syntax is surface VSO, but the most likely underlying word order is VOS, with a rule that normally moves all but the first phrase in a complex predicate to the right of the subject. The basic unit of syntax is the phrase, which has a grammatical particle indicating the phrase function preceding the lexical material. Modifiers usually follow the lexical head. The grammatical particles include markers of tense and aspect, and prepositions that indicate noun-phrase function, and may be tense-marked. The sentence subject is the only NP without an introductory preposition. Maori does not have a copular verb, and many sentences lack overt verbs. The basic syntax is illustrated in the following:

| | | | | |
|---|------------------|-------------------|-------------|------------|
| kei | raro | ngā | pukapuka | tawhito |
| at-PRES | <i>underside</i> | DEF.PL | <i>book</i> | <i>old</i> |
| a | Hone | i | | |
| DEEPL.APOSS | <i>John</i> | at. TNSNEUTRAL | | |
| t-ō-na | | moe-nga | | |
| DEESG-OPOSS-3SG | | <i>sleep-NOML</i> | | |
| 'John's old books are under his bed' | | | | |
| ka | tope | a | Wāka | |
| FUT | <i>chop</i> | PERS ART | <i>Wāka</i> | |
| i | te | rākau | rimu | a |
| ACC | DEESG | <i>tree</i> | <i>rimu</i> | at.FUT |
| te | Rāhina | | | |
| DEESG | <i>Monday</i> | | | |
| 'Wāka will chop down the rimu tree on Monday' | | | | |

The pronoun system distinguishes singular, dual, and plural, and in the first person, inclusive and exclusive. Number is marked on determiners, not nouns (except for a handful of personal nouns), the singular usually with an initial *t-* and the plural with Ø (e.g., *tētahi* – sg, *ētahi* – pl) and there is a special determiner for proper names. There is a very complex system for the expression of ownership, including a distinction akin to alienable/inalienable (marked by *a-* vs. *o-*forms). Most lexical items can serve without

morphological change in noun phrases, verb phrases, or as modifiers. In narratives, the passive is used for most event sentences, which has given rise to much debate about ergative vs. accusative syntax. A typologically unusual feature is that lexical modifiers of passive verbs take a passive suffix in agreement with the verb. The direct object is not integrated into the grammatical system of Maori in the way that would be expected if the language was originally accusative: for instance it cannot normally be relativized on directly. This is probably one remnant of a former ergative syntax.

Bibliography

- Bauer W A (1993). *Descriptive grammar series: Maori*. London: Routledge.
- Bauer W A (1997). *The Reed reference grammar of Māori*. Auckland: Reed Books.
- Benton R & Benton N (2000). 'RLS in Aotearoa/New Zealand 1989–1999.' In Fishman J A (ed.) *Can threatened languages be saved? Reversing language shift, revisited: a 21st century perspective*. Clevedon: Multilingual Matters Ltd. 423–450.
- Biggs B (1961). 'The structure of New Zealand Māori.' *Anthropological Linguistics* 3, 1–54.
- Biggs B (1973). *Let's learn Maori: a guide to the study of the Maori language* (2nd edn.). Wellington: A. H. & A. W. Reed.
- Hale K (1968). Review of P. W. Hohepa, *Memoir 20, International Journal of American Linguistics: A profile-generative grammar of Maori*. *Journal of the Polynesian Society* 77, 83–99.
- Harlow R B (2001). *A Māori reference grammar*. Auckland: Pearson Education NZ Ltd.
- Head L (1989). *Making Maori sentences*. Auckland: Longman Paul.
- Hohepa P W (1967). *Memoir 20, International Journal of American Linguistics: a profile-generative grammar of Maori*. Waverley, Baltimore: Indiana University Publications in Anthropology and Linguistics.
- King J (2001). 'Te kōhanga reo: Māori language revitalization.' In Hinton L & Hale K (eds.) *The green book of language revitalization in practice*. San Diego: Academic Press. 118–128.

Mapudungan

F Zúñiga, University of Zurich and University of Leipzig and Centro de Estudios Públicos, Zurich, Switzerland

© 2006 Elsevier Ltd. All rights reserved.

The dialects of Mapudungu(n), less deviant from one another than the dialects of English, are spoken by the Mapuche of south-central Chile and central Argentina. Current conservative estimates place the number of fluent speakers at one-third of the almost 1 000 000 ethnic Mapuche; more than 90% are in Chile, of which more than 40% are in or around Santiago and only 30% live in the traditional Mapuche territory. The main present-day dialects are (1) Mapudungun proper or Central Mapudungun, in south-central Chile, and (2) Pehuenche, to the east of the former. Further dialects such as Argentinean Ranquel and Chilean Picunche and Huilliche are either obsolescent or extinct. Other names for the language are Araucanian (from *araucanos*, the ethnonym used by the Spaniards; present-day Mapuche avoid using this term), Mapuchedungun, and (Re)chedungun. Several genetic affiliations have been proposed, not only with languages spoken in the south of the continent such as Kawésqar (Qawaskar) and

Yaghan (Yámana) but also with language families as distant as Arawak, Carib, and Mayan; to date, none of these proposals has been convincingly substantiated. The first grammar dates from the early 17th century. Written texts began to appear in the early 1900s and have become more numerous only at the end of the 20th century.

The phoneme inventory is simpler than the ones found in neighboring languages: the vowels are /a, e, i, o, u, i/ (where /i/ <ü> is unrounded high central when stressed and close to a schwa when unstressed). The glides are palatal /j/ <y>, labiovelar /w/, and velar /ɰ/ <g>. The consonants are the voiceless unaspirated noncontinuant /p, t̪<t>, t, ɬ<tr>, c <ch>, k/ (where /c/ is alveolopalatal), the voiceless fricatives /f, θ <d>, s <s, sh>/, the nasals /m, ñ <n>, n, ñ, ŋ <ng>/, and the liquids /l̪ <ɭ>, l, ʎ <ll>, ɭ<r>/. The dental series /t̪, ñ, l̪/ contrasts with the alveolar one /t, n, l/ only in highly conservative speech; most speakers have an alveolar series only. Pehuenche has voiced fricatives [v] and [ð] instead of [f] and [θ]. Primary stress can be largely predicted from syllable structure (it tends to fall on the penultimate mora) – with some exceptions, as in a number of disyllabic adverbs whose stress is lexically assigned to the ultima. Because there is no universally accepted orthographic convention,

it is not uncommon to find some variation in the literature and deviant spellings such as <z> instead of <d> (for /θ/), <g> instead of <ng> (for /ŋ/), and <(t)x> instead of <tr> (for /t/) in recent texts.

Nominal morphology is simple and almost exclusively derivational; there is neither gender nor case, adjectives take the suffix *-ke* in the nonsingular, and human nouns are marked with a preposed element *pu* in the plural. Compounding is highly productive, for example,

mapu-che
land-people
'people of the land'

Personal and possessive pronouns distinguish three persons (first, second, and third) and three numbers (singular, dual, and plural). Noun classes, possessive classes, classifiers, and alienable/inalienable possession are not marked overtly.

Agglutinative and predominantly suffixing, verbs are marked for mood (indicative *-i*, subjunctive *-l* and unmarked imperative), tense (future *-a* and unmarked nonfuture), evidentiality (reportative-mirative *-rke*), polarity (negative *-la* ~ *-no* ~ *-ki* and unmarked affirmative), directionals (e.g., cislocative *-pa*, translocative *-pu*, andative *-me*), voice (e.g., applicatives *-ñma* and *-(le)l*, reflexive *-u*, agentless passive *-nge*, causativizer *-m*), aspectuality (e.g., habitual *-ke*, progressive/resultative *-(kü)le*, progressive *-meke*, continuative *-ka*, telicizer *-tu*, ambulative *-(k)iaw*), modality in the broad sense (e.g., ruptured implicature *-fu*, immediate action *-fem*, sudden action *-rume*), and person. Person marking is intricate and can be described as following a direct/inverse pattern in that, when more than one argument is involved, the one ranking higher in the nominal hierarchy or primary argument (first- and second-person over third-person) and proximate over obviative on chiefly pragmatic grounds) is fully marked for person and number, whereas the lower or secondary argument is either underspecified or only implied. There are also forms that do not typically occur as predicates in main clauses; they do not encode the primary argument on the verb (but, with the exception of the *lu*-form, take an external possessive marker instead), function as verbal nouns, participles, and/or gerunds and end in *-n*, *-el*, *-lu*, *-yüm*, or *-am*. Many suffixes are transparently related to verb roots that still occur as such in the language (e.g., *meke* 'be busy,' *tu* 'take,' *kiaw* 'walk,' and *fem* 'do so'), and root serialization is used in order to express path of motion and other categories, for example,

rüngkü-kon-
jump-enter-
'jump in'

Reduplication of the root with *-tu* or *-nge* is used for the iterative, for example,

rüngkü-rüngkü-tu-
jump-jump-PRT
'jump repeatedly'

The incorporation of complex NPs is receding in the speech of younger urban Mapuche, but is productive in the speech of older rural speakers, for example,

ngilla-kurü-ka_l-ufisha-
buy-black-wool-sheep-
'buy wool of black sheep'

On the other hand, the incorporation of simple NPs into the verbal complex is still robust, as in

katrü-kachu-
cut-grass-
'cut grass'

Adjectives, wh-words, and numerals can all take verbal morphology.

Mapudungan is head-marking at both the NP level (possession) and the clause level. Unmarked utterances tend to be verb-initial and have at most one argument NP. There are both prepositions and postpositions, and adpositional phrases tend to follow the predicate and its arguments. Almost all marked word-order patterns are attested, and all can be elicited; an interplay of direct/inverse marking and position governs the interpretation of NPs as agentive or patientive. There do not seem to exist overt topic or focus markers, but a number of particles are used in questions or as intensifiers. Clause-linkage patterns resort to the gerunds previously mentioned or show coordination (e.g., with *ka* 'and' or *welu* 'but') or juxtaposition.

Although the lexicon of Mapudungan has visibly borrowed from Quechua and Spanish, the effects of contact on the morphology do not appear to be significant. The use of the verbal suffix *-fi* as more or less matching the use of the Spanish preposition *a* with direct objects and the higher frequency of AVO sentences in present-day Mapudungan are the most prominent contact-induced phenomena in the syntax.

Bibliography

- Augusta F J de (1903/1990). *Gramática mapuche bilingüe*. Santiago: Séneca.
Augusta F J de (1910/1991). *Lecturas araucanas*. Temuco: Kushe.
Augusta F J de (1916/1991). *Diccionario araucano*. Santiago: Cerro Manquehue.
Catrileo M (1995). *Diccionario lingüístico-etnográfico de la lengua mapuche: mapudungan – español – inglés*. Santiago: Andrés Bello.

Fernández Garay A (2001). *Ranquel – español/español – ranquel: diccionario de una variedad mapuche de La Pampa (Argentina)*. Leiden: CNWS, Universiteit te Leiden.

Harmelink B (1996). *Manual de aprendizaje del idioma mapuche: aspectos morfológicos y sintácticos*. Temuco: Universidad de la Frontera.

Salas A (1992). *El mapuche o araucano: fonología, gramática y antología de cuentos*. Madrid: Editorial Mapfre.

Smeets I (1989). *A Mapuche grammar*. Ph.D. diss., Universiteit te Leiden.

Zúñiga F (2000). *Mapudungun*. Munich: Lincom Europa.

Marathi

K Wali, Paris, France

© 2006 Elsevier Ltd. All rights reserved.

Marathi, a member of the Indo-Aryan subbranch of the Indo-Iranian branch of the Indo-European languages, is the official language of Maharashtra state of India. It is spoken by nearly 96 million people, according to the most recent census. The major dialects of Marathi are Deshi, spoken around Pune, Varhadī and Nagpuri, spoken around Nagpur, and Kokni in the coastal region. Marathi has been much influenced by Dravidian Kannada and Telugu spoken in the southern vicinity of Maharashtra (Bloch, 1920; Southworth, 1971).

History

Marathi is a direct descendent of Maharashtri, a prakrit language derived from Sanskrit. The earliest reference to spoken Marathi is found in *kuvalaymala* written in the 8th century by Udyotansuri. The earliest written Marathi is found in the 10th-century inscriptions at Shravan belgola. The earliest literary text is considered to be *Viveksindhu* of Mukundraja (1199 C.E.) A high literary form of Marathi appears around the 13th century in *Jnaneshawari*, a commentary on the *Bhagavad Gita* (Master, 1964).

Script

Modern Marathi script, called *balbodhi*, is based on the Sanskrit Devnagari script, with certain modifications. Unlike English, Devnagari is alphasyllabic. It uses certain diacritics for vowels when combined with consonants. The diacritics distinguish long and short vowels. There is also a special system to denote consonant clusters. There also is an alternative cursive script, called *modii*, which was introduced by Hemadpant around the 17th century and was used in official documents for some time.

Phonology

The traditional Marathi alphabetic chart lists 16 vowels and 36 consonants based on Sanskrit. Today, many of these alphabets are obsolete. Modern Marathi has 8 basic vowels and 34 consonants, including two semivowels. Tables 1 and 2 indicate the vocalic and consonantal charts and their respective features.

A salient feature of consonants is the distinction between affricates and palatals. The distinction is neutralized before y and i. The origin of affricates is controversial, because they are not found in Sanskrit.

Suprasegmentals

Length:

Vocalic length is mostly predictable. With the exception of ə, the last vowel of a word is long unless the vowels are followed by a combination of consonants such as *nt*, *tr*, *kt*. (iii) The length is phonemic in *i*, *u*.

Nasal vowels:

Use of nasal vowels as independent entities varies from speaker to speaker. They are found in certain adverbs, nouns, and plural nouns in the context of case and postpositions. They are phonemic in certain dialects.

Accent:

Marathi is said to have a stress accent. Length, pitch, and sonority play a role in determining the loudest accent. (For details, see Kelkar, 1958, 1997; Pandharipande, 1997.)

Table 1 Vowels

| | Front | Central | Back |
|---------|-------|---------|------|
| High | i | | u |
| Mid | e | | o |
| Mid low | ai/æ | ə | au/ɔ |
| Low | | a | |

Salient features: The qualitative difference between ə and a is not precise. ə can be extra short and silent. e and o occur in all positions. æ and ɔ are found mostly in borrowings from English.

Table 2 Consonants

| | <i>Labial</i> | <i>Dental</i> | <i>Retroflex</i> | <i>Alveolar</i> | <i>Alveopalatal</i> | <i>Velar</i> | <i>Glottal</i> |
|-------------------|---------------|---------------|------------------|-----------------|---------------------|--------------|----------------|
| Stops | | | | | | | |
| vcl.unasp | p | t | ʈ | | | k | |
| vcl.asp. | ph | th | ʈʰ | | | kh | |
| vcd.unasp | b | d | ɖ | | | g | |
| vcd.asp | bh | dh | ɖʱ | | | gh | |
| Affricates | | | | | | | |
| vcl.unasp | | | | c | č | | |
| vcl.asp. | | | | | čh | | |
| vcd.unasp. | | | | j | ǰ | | |
| vcd.asp. | | | | jʰ | ǰʰ | | |
| Nasals | m | n | ɳ | | | | |
| Laterals | | l | ɭ | | | | |
| Trill | | | | r | | | |
| Fricatives | | s | ʂ | | | | h |
| Semivowels | | v/w | | | y | | |

Salient features: *h* is a voiceless aspirate after voiceless stops and voiced aspirate in other positions. *s* becomes retroflex before a retroflex consonant. *v* vacillates between bilabial and labiodental position. It becomes a voiced lenis labiodental spirant in the initial position.

Morphology

Both animate and inanimate nouns exhibit two numbers – singular and plural – and three genders – masculine, feminine, and neuter. Marathi is a split ergative language. The subject is marked nominative with the exception of (i) transitive verbs in the perfective, (ii) obligative subjunctive, and (iii) dative verbs (Wali, 2004). The subject is marked ergative in (i and ii) and dative in (iii). In all these constructions, the verb agrees with the unmarked nounphrase, which may be a direct object or a theme. The verb shows neutral agreement if both subject and object have overt case. A salient feature of ergative system is seen in the pronominals. The first and second person pronouns are not overtly marked for ergative case and still show an ergative agreement pattern. What is more interesting is that the second person shows agreement for both nominative object and ergative pronoun though it is marked nominative (1). (See Wali, 2004.)

- (1) tu sāmāya ghas-l- ya-s.
 you- lamps-NOM- wash-PERF-
 NOM 3FPL 3FPL-2SG

‘You washed the lamps.’

Syntax

Word Order

The standard word order is subject object and verb, that is, SOV, in all constructions including interrogatives. The order is variable with certain restrictions. Most adjectives precede and agree with nouns. Adverbs precede the verb. Some adverbs agree with the verb.

Passivization

There are two types of passives: regular and capability. Although both are formed by adding *ja* ‘go’ to the verbal perfective, there is a difference. The former applies only to the transitives and allows demoted agent to be deleted. The latter operates across intransitives and transitives and does not allow the demoted agent to delete.

Subordination

The subordinate clause may be finite or nonfinite. It may precede or follow the main clause. Adverbial and relative clauses are correlative type. The latter allow deletion of head and correlative nouns and show a considerable range of word order variation (Wali, 1982). They exhibit a reduced participial form in all tenses. They also show a rare pattern of multiple headed relatives. These do not allow participial reduction.

Notion of Subject

Agreement is not a criterion for a subject because of the complexity noted in the morphology. It is necessary to resort to other grammatical rules such as reflexivization and participle reduction to determine the subject. Both nominative and ergative subjects undergo the same rules. However, the status of dative and passive subjects is enigmatic since they obey only some of these rules. In fact, the rule criterion leads to the conclusion that these two construction may have two subjects contrary to the traditional notion of there being a single subject in a sentence (see Wali, 2004).

Bibliography

- Bloch J M (1920). *La formation de la langue Marathe*. Paris: Champion.
- Master A (1964). *A grammar of old Marathi*. Oxford: Clarendon Press.
- Kelkar A R (1997). *Language in a semiotic perspective*. Pune: Shabda Sarswati Prakashan.
- Kelkar A R (1958). *Phonology and morphology of Marathi*. Ph.D. dissertation. Ithaca: Cornell.
- Pandharipande R V (1997). *Marathi*. London and New York: Routledge.

- Southworth F C (1971). 'Detecting prior creolization: An analysis of the historical origins of Marathi.' In Hymes D (ed.) *Pidginization and creolization of languages*. Cambridge: Cambridge University Press.
- Wali K (1982). 'Marathi correlatives: A conspectus.' In Pachori (ed.) *South Asian Review*, vol. 6:3. Florida: University of North Florida. 78–88.
- Wali K (2004). 'Non nominative subjects in Marathi'. In Peri B & Subbarao K V (eds.) *Non-nominative subjects*. Typological studies in language, vol. 61. Amsterdam: John Benjamins. 223–252.

Mayan Languages

J M Maxwell, Tulane University, New Orleans, LA, USA

© 2006 Elsevier Ltd. All rights reserved.

The Mayan language family traditionally stretched from what is now northern El Salvador and Honduras, through Guatemala and Belize, and up to the southern states of Mexico, including Chiapas, Quintana Roo, Campeche, Yucatán, and part of La Huasteca. Today the family is more dispersed due to out-migration. Large colonies of Mayan speakers can be found in Los Angeles and other California communities, Arizona, Texas, and Florida. Most linguistic descriptions recognize 31 Mayan languages, including the extinct Chikomulselteko (Chicomuceltec). Most historical linguists posit the Maya homeland as the Cuchumatán peaks of Guatemala, the area with the greatest linguistic diversity today (Kaufman, 1976; Campbell, 1977; Fox, 1978; Campbell and Kaufman, 1983, 1985). The model of diversification correlates phonological, morphological, and syntactic changes with a least-moves model of out-migration, seeking confirmation in the archaeological method. Based on these reconstructions, Proto-Maya, the mother language from which the modern diversity springs, would have been spoken approximately 41 000 years ago. People began to migrate outward, sharing innovations as they moved.

The family eventually split into four divisions. (Note that many of the names of Mayan languages have a variety of spellings. These spellings reflect not only the writing traditions of various authors (English, Hispanic, Mayan), but also their political orientation. In Guatemala, particularly, Mayans have fought for and won official recognition of their own orthographies. In Chiapas, Mayan educators have

also elected non-Spanish-based spelling systems. In Yucatán, however, a long tradition of literacy in Maya Yucateco has militated against changing orthographies. The spellings used in this article reflect the local practice.)

1. Wasteko, composed today only of Wasteko (Huasteco).
2. Yucatecan, composed of Maya Yucateco (Yucatán Maya), Mopan (Mopán Maya), Itzaj (Itzá), and Lakantun (Lacandán).
3. Western Division, broken into two branches: Ch'olan and Q'anjob'alan. The Ch'olan branch in turn has two subgroups, Ch'olan Proper, consisting of Chontal, Ch'ol, and Ch'orti' (Chortí), and Tzeltalan, consisting of Tzotzil and Tzeltal. The Q'anjob'al branch has two subgroups, Chujean, consisting of Tojolab'al and Chuj, and Q'anjob'alan Proper, consisting of Q'anjob'al (Eastern Kanjobal), Akateko (Western Kanjobal), Popti' (formerly Jakalteko (Jacalteco)), and Mocho' (Mochó).
4. Eastern Division, which is subdivided into the Mamean and K'iche'an subgroups. The Mamean branch is broken into Mam Proper, consisting of Mam and Teko (also called Tektiteko (Tectiteco)), and Ixilan, which includes Ixil (Nebaj Ixil) and Awakateko (Aguacateco). The K'iche'an branch includes the outliers Uspanteko (Uspanteco) and Q'eqchi' (Kekchí) and two major subdivisions, K'iche'an Proper, consisting of K'iche' (Quiché), Achi', Kaqchikel (Central Cakchiquel), Tz'utujil (Tzutujil), Sakapulteko (Sacapulteco), and Sipakapense (Sipacapense), and Poqom, consisting of Poqomchi' (Pocomchí) and Poqomam (Pocomam).

Within subgroups there is a high degree of mutually intelligibility and multilingualism (particularly

evident in market contexts), which blurs language and dialect boundaries. Often the divisions between language groups are determined more by political divisions and historical identities than by isoglosses. Rivalry between families in Aguacatán brought about the splintering of Awakateko, spawning a new ‘language,’ Chalchiteko, which won official recognition in Guatemala in 2003. Likewise, historic autonomy and a tradition of armed and political conflict between Q’umark’aj (the K’iche’ capital) and Rab’in’al (the Achi’ center) have created localized identities, which override mutual intelligibility in determining the language boundaries between the two groups. Residents of San Miguel Acatán and San Rafael La Independencia have traditionally considered themselves Q’anjob’al speakers, but the official recognition of Akateko as a language, with its own representation in the Academy of Mayan Languages of Guatemala, has served to accentuate linguistic differences and has discouraged use of educational materials, more widely available in Q’anjob’al. On the other hand, Mam, one of the four largest Guatemalan Mayan languages, and Chuj, spoken in northwestern Guatemala, both have deep internal dialectal splits. Dialects may differ in more than 20% of their core vocabulary, undergo different syntactic processes, and allow different sentential word orders, yet these languages maintain a shared identity.

Estimates of number of speakers are also highly political. Despite official rhetoric praising the ethnic

richness of their countries, both Guatemala and Mexico have traditionally promoted assimilation to a national identity that is indigenous only ancestrally. (Also, El Salvador does not recognize any modern Maya as traditional ethnicities, although it does again host Mayan populations displaced by the genocidal war in Guatemala, 1960–1995. Honduran populations until recently were counted as Spanish speaking, although in the north there were ethnically Ch’orti’ peoples; in the 2001, some Honduran rural schools began limited bilingual education, although without materials.) Leopoldo Tzian (1994) points out that official governmental censuses in Guatemala consistently underestimate the number of Mayas compared to surveys done by linguists, by international development agencies, and by health workers. **Table 1** gives population figures for Guatemala: official census figures for the Mayan population (note the difference between ethnically identified Maya and those who speak their mother tongue), Tzian’s data, the figures of AJPOPAB’CHI’ (the Commission for the Officialization of the Indigenous Languages of Guatemala), and those of the Ministry of Education Survey for 2003. **Table 2** contains population figures for Mexico, showing the official government figures (Instituto Nacional de Estadística, Geografía e Informática, 2000) and those of the Summer Institute in Linguistics (published 2004) with the date of the survey in parentheses. The first label under the rubric ‘language’ gives the traditional name for the

Table 1 Population figures

| <i>Language</i> | <i>Ethnic count, 2002</i> | <i>Speaker count, 2002</i> | <i>Tzian (1994)</i> | <i>Ajpopab’achi’, 1998</i> | <i>Ministry of Education, 2003</i> |
|-----------------|---------------------------|----------------------------|---------------------|----------------------------|------------------------------------|
| K’iche’ | 1 270 953 | 890 596 | 1 842 115 | 647 624 | 922 378 |
| Q’eqchi’ | 852 012 | 716 101 | 711 523 | 473 749 | 726 723 |
| Mam | 617 171 | 477 717 | 1 094 926 | 345 548 | 519 664 |
| Kaqchikel | 832 968 | 444 954 | 1 002 790 | 343 038 | 475 889 |
| Q’anjob’al | 159 030 | 139 830 | 205 670 | 75 155 | 99 211 |
| Poqomchi’ | 114 423 | 92 941 | 259 168 | 94 714 | 69 716 |
| Ixil | 95 315 | 95 315 | 130 773 | 47 902 | 69 137 |
| Achi | 105 992 | 82 640 | n/a | 15 617 | 51 593 |
| Tz’utujil | 78 498 | 63 237 | 156 333 | 57 080 | 47 669 |
| Chuj | 64 438 | 59 048 | 85 002 | 50 000 | 38 253 |
| Popti’ | 47 024 | 34 038 | 83 814 | 39 635 | 38 350 |
| Akateko | 39 370 | 16 562 | 39 826 | 40 991 | 5572 |
| Ch’orti’ | 46 833 | 11 734 | 74 600 | 27 097 | 9105 |
| Poqomam | 42 009 | 11 273 | 127 206 | 46 515 | 9548 |
| Awakateko | 11 068 | 9613 | 34 476 | 18 572 | 16 272 |
| Sakapulteko | 9763 | 6973 | 42 204 | 3 033 | 3940 |
| Sipakapense | 10 652 | 5687 | 5944 | 4 409 | 6344 |
| Uspanteko | 7494 | 3971 | 21 399 | 12 402 | 1231 |
| Mopan | 2891 | 2456 | 13 077 | 8500 | 468 |
| Itzaj | 1983 | 1094 | 1783 | 650 | 123 |
| Teko | 2077 | 1144 | 4755 | 4895 | 1241 |
| Chalchiteko | n/a | n/a | n/a | n/a | n/a |

language/ethnic group, used in most academic publications and in official documents prior to 2000, the second is the indigenous autodenomination.

Grammatical Characteristics

The Mayan Languages share many important characteristics, among these are ergativity, positionals, directional particles, and noun and numeral classifiers. These categories are developed in different ways in the various languages.

Ergativity

Ergative languages mark the relationship between the verb and its arguments with inflections that treat the subjects of intransitives and objects of transitive verbs as one category (marked by absolutive pronouns) and the subjects of transitives and possessors of nouns as a separate category (marked by ergative pronouns). Most of the Mayan languages show this system, with variations in subparts of the grammar, in which a nominative–accusative agreement pattern (like that Indo-European languages) surfaces. Such systems are referred to as split-ergative. Ch'orti' has a split-ergative system, with the change being triggered by subordination. In addition, Ch'orti' has a third

pronominal set, which serves as prefixed subject markers of incomplete intransitive verbs. Table 3 shows sample verbs in Kaqchikel with subject pronouns in bold type and object pronouns in italics. Note the homology of intransitive subjects and transitive objects.

Positionals

Positionals are a special word class in Mayan languages, so-called because many denote positions such as 'standing,' 'lying prone,' and 'stuck crosswise in an opening.' However, some simply name conditions or states, such as 'wet,' 'naked,' and 'round.' Words that belong to this class have special derivational characteristics. The roots are inflected to form two or three types of nonverbal predicates (adjectives), intransitive verbs, and transitive verbs. Table 4 shows examples from Mam.

Some languages form reduplicated adjectives from the positional roots, for example, in Chuj *nhojanho-jan* 'walk fluffily, like a shaggy sheep' and *linganlingan* 'be hanging.' Kaqchikel (Table 5), Tz'utujil, and K'iche' copy the vowel of the root and the first consonant and then add a suffix for singular or plural agreement to form adjectives from positional roots.

Table 2 Mexican population figures

| Language | Government census figures, 2000 ^a | Summer Institute in Linguistics |
|----------------------------|--|---------------------------------|
| Tzeltal, K'op | 547 000 | 300 000 (1993) |
| Tzotzil, B'atzil K'op | 514 000 | 225 000 (1990) |
| Mocho | 500 | >1000 (1993) |
| Lakantun, Hach T'an | 130 | |
| Tojolab'al, Tojowinik Otik | 74 000 | |
| Chontal | 72 000 | |
| Ch'ol | 274 000 | |
| Yucateco | 1 490 000 | |

^aThe Mexican census also lists the numbers of speakers of 'Guatemalan' languages now resident in Mexico: Chuj 3900, Jakateko (Popti', Ab'xab'al) 1300, Q'eqchi' 1700, K'iche' 640, Kaqchikel 610, Ixil 310, Awakateko 60, Teko 50.

Table 3 Sample verbs in Kaqchikel

| First-person plural | Gloss | Second-person plural | Gloss |
|---------------------|-------------------|----------------------|-------------------|
| xojwa' | 'we ate' | xixwa' | 'y'all ate' |
| xixqaq'etej | 'we hugged y'all' | xoqiq'etej | 'y'all hugged us' |

Table 4 Sample positionals in Mam

| State | Gloss | Intransitive | Gloss | Transitive | Gloss |
|---------|------------|--------------|---------------|------------|-------------------------|
| wa'li | 'standing' | wa'ee | 's/he stands' | twa'b'in | 's/he stood her/him up' |
| xhjewli | 'twisted' | xhjewee' | 'it twists' | txhjebw'in | 'she/he twists it' |

Table 5 Reduplicated adjectives in Kaqchikel

| <i>State</i> | <i>Gloss</i> | <i>Adjective</i> | <i>Gloss</i> |
|--------------------|--|--------------------|------------------------------|
| setesik kotokik | 'round, singular' 'crooked, singular' | setesäq kotokäq | 'round, plural' 'crooked' |

Table 6 Poqomam directionals

| <i>Directional</i> | <i>Gloss</i> | <i>Intransitive verb</i> | <i>Gloss</i> | <i>Phrasal use</i> | <i>Gloss</i> |
|--------------------|--------------|--------------------------|----------------|--------------------|--------------------------|
| ala | 'out' | -il- | 'leave' | xa'ila ala | 'we go out, we leave' |
| aka | 'in' | -ok- | 'enter' | xah'oka aka | 'we go in, we enter' |
| koon | 'stay' | -kahn- | 'stay, remain' | xahkahna koon | 'we stay here' |
| pa | 'thither' | -pan- | 'arrive there' | xahpana pa | 'we arrive there' |
| qa | 'down' | -qaj- | 'descend' | xahqaja qa | 'we go down, we descend' |

Table 7 Noun classifiers in Popti'

| <i>Classifier</i> | <i>Objects in the class</i> | <i>Classifier</i> | <i>Objects in the class</i> |
|-------------------|-----------------------------------|-------------------|------------------------------|
| komam | Male supernaturals, diseases | komi' | Female supernaturals |
| ya' | Adult person | ho' | Young men |
| xo' | Young women | naj | Male, unknown, not respected |
| ix | Female, unknown, not respected | unin | Human baby |
| no' | Animals (other than the dog) | metx | Dog |
| te' | General plants and their products | ixim | Grains |
| tx'al | Cotton or synthetic thread | tx'anh | Fiber, string |
| q'ap | Cloth | tx'otx' | Earth, earthenware |
| ch'en | Metal, rock, mineral | atz'am | Salt |
| ha' | Water, liquid | q'a' | Fire |

Directional Particles

These particles, usually variants of intransitive verbs of motion, serve as a complement to main verbs. They may indicate actual movement of the actor or action, or they may add aspectual information. In Mam, transitive verbs almost always cooccur with a directional complement (see Table 6 for examples in Poqomam). Verb phrases in Yucatec, however, now rely on conjunction rather than complementation.

Noun Classifiers

These particles precede the nouns they modify and ascribe some property, social or material, to the noun. In the Q'anjob'alan group, noun classifiers are highly exploited by the grammar. They serve as definite articles and as pronouns. (See Table 7 for examples in Popti'.) In the neighboring Mamean languages, the system is more attenuated. In K'iche'an languages, classifiers are used more as titles before names than as classifiers. In Yucatec, only morphological vestiges appear in names for a few plants and animals.

Numeral Classifiers

These may be of two types of numerical classifiers: suffixal, marking what kind of entity is being counted (Table 8), and independent, showing how the object counted is measured (Table 9). The suffixal type distinguishes three classes in Q'anjob'alan languages: people, animals, and other. Other Mayan languages have only trace suffixes, sometimes invariant in form.

Vocabulary

Mayan languages have borrowed words from many languages, including Nahuatl (Náhuatl) (*masat* 'deer,' *tinamit* 'town,' in Kaqchikel), Spanish (*mexa* 'table,' *kaxtilanh winakhin* 'I'm a Spaniard, cock's crow,' in Chuj), and English (*tab'ana klik pa ruwi' ruk'in ri maus* 'click on it with the mouse,' in Kaqchikel). They have also lent many words, for example, English *hurricane* < Kaqchikel *juraqän* '(lit.) one leg' and Spanish *makuy* < *majkuy* 'an herb.' New words are constantly developed with the contact of cultures and the implementation of new educational curricula.

Table 8 Popti' numeral classifier suffixes

| Number root | Gloss | Root with suffix | Gloss |
|-------------|--------|------------------|---------------|
| kanh- | 'four' | kanhwanh | 'four people' |
| waj- | 'six' | wajk'onh | 'six animals' |
| b'alunh- | 'nine' | b'alunhe' | 'nine things' |

Table 9 Te'utujil measure words^a

| Measure | Gloss | Number root | | Combined form | Gloss |
|---------|---------------------|-------------|---------|-----------------|---------------------|
| mooq' | 'fistful' | ox- | 'three' | oxmooq' | 'three fistfuls' |
| quum | 'sip' | kaj- | 'four' | kajquum | 'four sips' |
| tz'uur | 'drop' | juu- | 'one' | juutz'uur | 'one drop' |
| seel | 'slice, layer' | ka'- | 'two' | ka'seel | 'two slices' |
| peer | 'plane surface' | waq- | 'six' | waqpeer tz'alam | 'six planed boards' |
| raab' | 'long, cylindrical' | wuq- | 'seven' | wuqraab' kolo' | 'seven ropes' |

^aNote that the measure word or classifier serves as the base. The number is prefixed in an abbreviated combinatorial form.

In Guatemala, the Academia de las Lenguas Mayas de Guatemala, an semi-autonomous branch of the government, is authorized to promote and develop the national languages. In Mexico, the federal government provides bilingual educational support and is supplemented by the efforts of the Academia de La Lengua Maya in Yucatán, Campeche, and Quintana Roo and by Sna Tz'ib'alom, the independent writers' cooperative in Chiapas.

Bibliography

- AJPOPAB'CHI (1998). *Propuesta de modalidad de oficialización de los idiomas indígenas de Guatemala*. Comisión de Oficialización de los Idiomas Indígenas de Guatemala. Guatemala City, Guatemala: Editorial Nojib'sa.
- Campbell L (1977). *University of California publications in linguistics 81: Quichean linguistic prehistory*. Berkeley, CA: University of California Press.
- Campbell L & Kaufman T (1983). 'Mesoamerican historical linguistics and distant genetic relationships: getting it straight.' *American Anthropologist* 85, 362–372.

Campbell L & Kaufman T (1985). 'Mesoamerican linguistics: where are we now?' *Annual Review of Anthropology* 14, 187–198.

England N C (1992). *La autonomía de los idiomas mayas: historia e identidad*. Guatemala City, Guatemala: Cholsamaj.

Fox J (1978). *Proto-Mayan accent, morpheme structure, conditions and velar innovations*. Ph.D. diss., University of Chicago.

Instituto Nacional de Estadística, Geografía e Informática (2000). *Censo nacional*. Guatemala City, Guatemala: Cholsamaj.

Kaufman T (1976). 'Archaeological and linguistics correlations in Mayaland and associated areas of Meso-America.' *World Archaeology* 8, 101–118.

Tzian L (1994). *Kajlab'aliil Maya'iib' xuq mu'siib': ri ub'an-tajjik Iximuleew*. Mayas y ladinos en cifras: el caso de Guatemala. Guatemala City, Guatemala: Cholsamaj.

Relevant Website

<http://www.sil.org> – Ethnologue.

Michif

P Bakker, Aarhus University, Aarhus, Denmark

© 2006 Elsevier Ltd. All rights reserved.

Michif (/mitʃif/; Mitchif, Mitif, erroneously also spelled Métchif) is a mixed language spoken by some of the Métis in Canada and the United States. The Métis are a separate ethnic group, who came into being as a

result of European men marrying Amerindian women. Around the year 2000, the number of speakers had dwindled to fewer than 1000, in scattered locations mostly in Saskatchewan, Manitoba (Canada), North Dakota, and Montana (United States). All speakers are elderly. The language is highly endangered.

Mixed European-Amerindian marital unions took place from the first days of contact in New France.

In the early period, the children of these relationships were typically raised by the mother as aboriginal children. In the early 1800s, the children of mixed parentage increasingly identified as separate from Europeans and from Amerindians, which was strengthened by political developments such as free trade, in which many such people were involved (Dickason, 1985). They called themselves *La Nouvelle Nation*, and also Mitif, Métis, Halfbreed, and similar terms. The name of the language is regularly derived from the term Mitif, which was the name for a person of mixed parentage in New France in the early 1600s.

The Métis combine aspects of Native cultures and French Canadian culture. Traditionally, the Métis had a very diverse lifestyle throughout the 19th century. Many were farmers, hunters, traders, and/or craftsmen. The hunters were especially famous for their massive buffalo hunts, until the near extinction of this animal in the late 1800s. Linguistically, the Métis were and are diverse as well, counting speakers of Ojibwe, Cree (both Algonquian), French, English (mostly in recognizable ethnic variants) as well as Michif, which combines, roughly, Cree verbs with French nouns. This is a typical example, from a fairy tale told by Norman Fleury in January 2004 (French elements in bold):

- (1) **sa** **fam,** **sa** **pramyer**
 3-POSS woman 3-POSS first-F
fam kii-wanih-eww
 woman PST-lose.AN-3.SUBJ.3OBJ
 ‘He had lost his wife, his first wife.’
- (2) eekwa kiihtwam kii-wiiw-eww
 then again PST-marry-3.SUBJ.3OBJ
 ‘And now he had remarried.’
- (3) kii-wiikim-eww onhin **la** **fam-a**
 PST-marry-3.SUBJ. this-OBV the woman-OBV
 3OBJ
 ‘He had married this woman.’
- (4) **sitenn** **moves** **fam,** **enn** **moves** **fam**
 it.was.F bad-F woman a-F bad woman
 ‘She was a bad woman. A bad woman.’
- (5) kii-machi-manitu-wi-w **awa** **la** **fam**
 PST-bad.spirit-BE-3SG this-AN the-F woman
 ‘She was a real devil, this woman.’
- (6) **pi** **ilave** **trwaa** **fiy** **ana**
 and she-had three girl this-AN.SG
 kii-ayaw-eww **la** **fam**
 PST-have-3SG the-F woman
 ‘And she had three daughters, this woman had.’

This fragment, using a unified orthography for the two components, illustrates some aspects of the mixture of the language. In this fragments all verbs are from Cree, except the two copulas *sitenn*

(F. *c'est une*) and *ilave* (F. *il/elle avait*). The nouns are all from French, including definite and indefinite articles and possessives (*la, lii, aen, enn*), the demonstratives from Cree (*awa, ana, onhin*), numerals from French (*trwaa, pramyer*), and conjunctions and adverbs from Cree (*kiihtwam, eekwa*) or French (*pi < F. puis*).

Sentential word order is free: (1) shows Object-Verb word order, (3) Verb-Object. Some preverbal modifiers can be separated from their nouns, notably numerals from the nouns, and in (6) a demonstrative from the noun. In morphology, French elements display French derivational and inflectional elements, whereas Cree elements combine with Cree morphology. The Cree noun *mushum* ‘grandfather’ gets Cree possessives and plural morphemes (*ni-mushum-ak* 1SG-grandfather-PL ‘my grandfathers’). There are a few exceptions, such as the nontopic or so-called obviative suffix as in *la fam-a* in (3) that is often added to French nouns, and borrowed or code-switched French and English verbs as in *gii-li-park-ii-naann* 1PST-MARKER-park-INF-1PL ‘we parked’ or *kii-li-move-ii-w* PST-MARKER-move-INF-3SG ‘she moved (away)’. This sentence switches languages four times: Cree-French-English-French-Cree.

The phoneme inventory of Michif is a combination of the inventories of the Métis variety of French (influenced by a Northern dialect of Cree) and South-eastern Plains Cree (influenced by Saulteaux Ojibwe). For instance, preaspirated consonants and nasal /i/ are only found in words of Cree etymology, and phonemes such as /r/ and /l/ are only found in the French part. Strikingly, the range of allophones for Cree and French phonemes differs: whereas voiced /b/ and unvoiced /p/ (and other stops and fricatives) are allophones in Cree between vowels, they are distinct phonemes in the French part. The Cree and French components also follow their own phonotactic patterns. Only stress patterns seem to have moved in the direction of Cree.

Michif combines Cree and French agreement: the (Algonquian) animacy of nouns is reflected in the gender inflection of demonstratives and verbs, and the (French) masculine-feminine distinction is shown in definite and indefinite articles and preverbal adjectives.

Michif is more complex than both of its components, having two parallel semantic, phonological, morphological, and syntactic systems. It combines the complexity of the Algonquian verb, with hundreds of forms, with the irregularities of French nominal derivation.

The language probably came into being in the early 19th century, parallel with the development of a new identity of mixed persons as a new ethnic group.

Michif is associated with descendants of the Métis buffalo hunters and their winter camps, and it is likely that these hunts played a role in the dissemination of the language.

Michif belongs to the small set of mixed or intertwined languages, of which a few dozen examples are known from all parts of the world. Whereas other such languages usually combine the grammatical system of one language with the lexicon of another, Michif seems to be almost unique in that it combines verbs from one language with nouns from another. Only the Nigerian language Igbo-Okrika seems to display a similar pattern (Igbo verbs, Ijo nouns).

Bibliography

- Bakker P (1997). "A language of our own." *The genesis of Michif – the mixed Cree-French language of the Canadian Métis*. New York: Oxford University Press.
- Dickason O P (1985). "From one nation" in the Northeast to "New Nation" in the Northwest: a look at the emergence of the Métis.' In Peterson J & Brown J (eds.) *The new peoples. Being and becoming Métis in North America*. Winnipeg: University of Manitoba Press. 19–36.
- Laverdure P & Allard I R (1983). *The Michif dictionary. Turtle Mountain Chippewa Cree*. Winnipeg: Pemmican Publications.

Misumalpan

A Constenla Umaña, University of Costa Rica, San José, Costa Rica

© 2006 Elsevier Ltd. All rights reserved.

Misumalpan is a Central American linguistic family with five members: Cacaopera, Matagalpa, Miskito, Sumo (Sumo Tawahka), and Ulwa. Cacaopera was spoken in eastern El Salvador and became extinct during the first half of the 20th century. Matagalpa, once diffused through the western portions of mid- and northern Nicaragua, and in the immediate adjacent zone of southern Honduras, became extinct at the end of the 19th century. Sumo is the language of 9000 people in northeastern Nicaragua (the area of the Waspuk and Bambana rivers) and about 500 people in the neighboring area of Honduras (along the middle Patuca River). Sumo has two dialects: Panamahka, which includes 80% of the speakers, and Tawahka. Ulwa, formerly widespread in south-eastern Nicaragua, is currently spoken by about 500 people in the area of the lower Grande de Matagalpa and Kuringwas rivers. Miskito is spoken in the Caribbean coast and neighboring lowlands from the Black River in Honduras (25 000 speakers) to the Pearl Lagoon in Nicaragua (100 000 speakers). The following dialects are mentioned: Mam (Honduras), Wangki, Tawira, Baldam, and Kabo (Nicaragua).

Ulwa and Sumo on the one hand, and Matagalpa and Cacaopera on the other, constitute two uncontroversial subgroups. Lexicostatistics show a closer relationship between them than that of either one to Miskito, so the family seems to be basically

divided into the latter and what we could call Western Misumalpan. The time depths according to glottochronology are Miskito/Western Misumalpan, 5800; Sumo–Ulwa/Matagalpa–Cacaopera, 5300; Matagalpa/Cacaopera, 1200; and Sumo/Ulwa, 880.

The only external relationship that has been proven by means of the comparative method is that with the Lencan family from Honduras and El Salvador. The split of the common ancestor would have taken place about 7200 years ago.

Many have thought Misumalpan to be related to Chibchan. This seems probable but has not yet been properly demonstrated.

The Misumalpan languages belong to the Lower Central America linguistic area, characterized by features such as SOV order, postpositions, preposed genitive, postposed numerals and adjectives, and contrasts between voiced and voiceless stops. Inside the area, they are part of a Northern Subarea characterized by features such as person inflection for possession in nouns and for agent and patient in verbs, predominance of accusative-nominative case systems, serial verb constructions, postposed or suffixed negation, and vowel length contrasts.

Bibliography

- Bertoglia M (1988). 'La lengua cacaopera: fonología y léxico.' Masters thesis, University of Costa Rica.
- Constenla A (1987). 'Elementos de fonología comparada de las lenguas misumalpas.' *Revista de Filología y Lingüística de la Universidad de Costa Rica* 13, 129–166.

- Constenla A (2002). 'Acerca de la relación genealógica de las lenguas lencas y las lenguas misumalpas.' *Revista de Filología y Lingüística de la Universidad de Costa Rica* 28, 189–205.
- Green T M (1999). 'A lexicographic Study of Ulwa.' Ph.D. dissertation, Massachusetts Institute of Technology.

- Norwood S (1997). *Gramática de la lengua Sumu*. Managua: Centro de Investigaciones y Documentación de la Zona Atlántica.
- Salamanca D (1988). 'Elementos de gramática del miskito.' Ph.D. dissertation, Massachusetts Institute of Technology.

Mixe-Zoquean Languages

R A Rhodes, University of California,
Berkeley, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

The Mixe-Zoquean languages are found in southern Mexico in the area of the Isthmus of Tehuantepec. The family is divided into two branches, Mixe languages in the Oaxacan highlands and the Zoquean languages, which are found on the Gulf of Mexico and in western Chiapas. The languages are:

Zoquean

- Gulf Zoque
 - Sierra Popoluca
 - Texistepec
 - Ayapa Zoque (Tabasco Zoque)
- Southern Zoque
 - Chimalapa
 - Copainalá
 - Francisco León
 - Zoque de Rayón

Mixean

- Veracruz Mixe
 - Oluteco (Oluta Popoluca)
 - Sayuleño (Sayula Popoluca)
- Highland Mixe
 - Western Mixe
 - Totontepec
 - Tlahuitoltepec
 - Eastern Mixe
 - Coatlán
 - Isthmus
 - Quetzaltepec
 - Juquila
 - Mazatlán

The Southern Zoque region of western Chiapas and the Highland Mixe region are dialect chains, such that the number of dialects and varieties is unevenly reported. The division given above represents varieties that show significant differentiation as measured by degree of mutual intelligibility (Grimes, 2004).

Historically, the ancestors of the Mixe-Zoqueans were the Olmecs (Campbell and Kaufman, 1988). This once controversial assertion was proved in the 1990s with the decipherment of Epi-Olmec hieroglyphics by Terry Kaufman and John Justeson, who have shown incontrovertably that the language represented is Zoquean (Kaufman and Justeson, 2003).

Phonology

The most salient phonological patterns of the Mixe-Zoquean languages are best understood through the phonological patterns of the protolanguage. Proto-Mixe-Zoquean had a six-vowel system, both long and short.

| | | |
|----------|----------|----------|
| <i>i</i> | <i>i</i> | <i>u</i> |
| <i>e</i> | <i>a</i> | <i>o</i> |

The consonantism was

| | | | | |
|----------|--------------|-----------|----------|----------|
| <i>p</i> | <i>t</i> | <i>ts</i> | <i>k</i> | <i>ʔ</i> |
| <i>m</i> | <i>n</i> | | | |
| | <i>s</i> | | | <i>h</i> |
| | (<i>l</i>) | | | |
| | <i>j</i> | <i>w</i> | | |

The basic shapes of proto-Mixe-Zoquean roots are:

| | | |
|--------|----------------|----------|
| CVC | <i>*tik</i> | 'house' |
| CVCC | <i>*ni:ʔh</i> | 'water' |
| CVCV | <i>*ni:wi</i> | 'chile' |
| CVCCV | <i>*tahpi</i> | 'hawk' |
| CVCVC | <i>*kiʔak</i> | 'sandal' |
| CVCCVC | <i>*pistik</i> | 'flea' |

The possible first consonants of these root internal clusters are very limited, *ʔ*, *h*, *s*, and nasals. There is one further pattern that exists only in verb roots. The (*ʔ*)C[sib] cluster allows only *p* and *k* as the medial consonant.

CV(*ʔ*)C[sib]- **heps-* 'shovel'

The phonotactics of proto-Mixe-Zoquean words allow a full array of clusters as a result of morphemic combination.

There were four phonological patterns in Proto-Mixe-Zoquean that leave significant reflexes in the modern languages. The first pattern featured metathesis of glottal stop to the front of a consonant cluster in word internal construction. When that metathesis crossed an obstruent, the obstruent was voiced. This is still an active process in some Zoquean languages, as shown in (1), and has left frozen traces in Mixean as in (2).

- (1) Sierra Popoluca (Zoque)
- (a) *ʔaʔniʃpa* 'I see him/her'
ʔan -*ʔiʃ* -*pa*
 1ERG -see -INCOMP
- (b) *tigiʔjpa* 'he has a house'
Ø -*tik* -*ʔiʔj* -*pa*
 3ABS -house -have -INCOMP
- (2) Sayuleño (Mixe)
- ʔigijp* 'he enters'
Ø -*ʔik* -*ʔij* -*p*
 3ABS -house -VERBALIZER -INCOMP

In the second pattern in syllable coda obstruents were augmented with a preceding *h*. The reflexes of this pattern yield allomorphies of morphemes with and without coda *hs* in both Zoquean and Mixean languages, although in most languages the patterns are radically restructured. In a few languages, including Sierra Popoluca and Oluta, the pattern was leveled away.

- (3) Francisco León Zoque
- (a) *petpa* 'he sweeps'
Ø -*peht* -*pa*
 3ABS -sweep -INCOMP
- (b) *pehtu* 'he swept'
Ø -*peht* -*u*
 3ABS -sweep -COMP

In Sayuleño (Mixe) the restructured pattern is one of *h* metathesis, as shown in (4).

- (4) Sayuleño (Mixe)
- niʔeʔhpip* 'he sees himself'
Ø -*ni* -*ʔeʔp* -*hi* -*p*
 3ABS -REFL -see -MIDDLE -INCOMP

The third pattern affected a class of roots containing V:ʔC. In construction with a consonant initial suffix the glottal was deleted. In construction with a vowel initial suffix the glottal was retained. This allomorphy has been leveled away in Zoquean.

- (5) preconsonantal prevocalic/preglide
- (a) Oluteco (Mixe)
- ʔimipa* 'he comes' *miʔn* 'come'
Ø -*ʔi:ʔn* -*pa* *ʔi:ʔn* -*i*
 3ABS -come -INCOMP come -IMP
- (b) Totontepec (Mixe)
- diʔe:tp* 'I sweep it'

| | | |
|--------------|---------------------|-------------|
| <i>di</i> | - <i>ʔe:ʔt</i> | - <i>p</i> |
| 3ERG | -sweep | -INCOMP |
| <i>diʔeʔ</i> | '(that) I sweep it' | |
| <i>di</i> | - <i>ʔe:ʔt</i> | - <i>I</i> |
| 3ERG | -sweep | -INCOMP_DEP |

The last pattern in the phonology of the protolanguage featured a limited height harmony in a few suffixes. The reflexes remain most clearly in the Zoquean languages.

- (6) Copainalá Zoque
- haʔn mini* 'he doesn't come'
haʔn j -*min* -*i*
 NEG 3ABS -come -NEG.INCOMP
- haʔn mjone* 'he doesn't wrap it'
haʔn j -*mon* -*i*
 NEG 3ABS -wrap -NEG.INCOMP

Much of Mixean has lost unstressed vowels. One effect is that vowels were lost from final syllables. One of the key phonological developments related to this vowel loss is that the Highland Mixe languages developed an umlaut that expanded the vowel systems in some dialects to three-height nine vowel systems with higher vowels triggered by suffixes that had high vowels.

- (7) Totontepec Mixe
- (a) normal
- ʃehp* 'I breathe' (< *ʔiʃehpa)
Ø -*ʃeh* -*p*
 1ABS-breathe-INCOMP
- raised
- nʒeh* '(that) I breathe' (< *ʔinʒehi)
n -*ʃeh* -*I*
 1ABS_DEP-breathe-INCOMP_DEP
- (b) normal
- tʉ:mp* 'I work' (< *ʔiʔtʉ:mpa)
Ø -*tʉ:n* -*p*
 1ABS-work-INCOMP
- raised
- ndun* '(that) I work' (< *ʔinʔtʉ:ni)
n -*tʉ:n* -*I*
 1ABS_DEP-work-INCOMP_DEP

Syllable coda *ws* in Zoquean are nasalized to *ŋ*. In all varieties the alternation is at least partially leveled leading to a marginal contrast.

- (8) Copainalá Zoque
- tsinba* 'he bathes' *tsiŋ* 'he bathed'
Ø -*tsiŋ* -*pa* *Ø* -*tsiŋ* -*u* (-*u* < **wi*)
 3ABS -bathe -INCOMP 3ABS -bathe -COMP
- tsiwi* 'bathe!'
tsiŋ -*i*
 bathe -IMP
 (cf. Say. *chiwjahpa* 'they bathe')

There are two other phonological developments in the Zoquean branch that are typologically notable.

Most varieties of Zoque have lost vowels from initial person markers. The result is those that contain *i* historically synchronically contain *j*, which metathesizes to the onset final position. This metathesis generalizes to other *j*C clusters.

(9) Copainalá Zoque

| | | | |
|-------------------|-------------------------|------------|--------------|
| <i>mbjopjamih</i> | 'you (sg.) are running' | | |
| <i>nj</i> | <i>-poj</i> | <i>-pa</i> | = <i>mih</i> |
| 2ABS | -run | -INCOMP | = 2 |

Finally, Texistepec has undergone a typologically unusual sound change, denasalizing root initial nasals in nonnasal environments, e.g., *bok* 'maize,' *d'ew* 'chile,' cf. Copainalá *mok*, *niwi*.

Morphology

The Mixe-Zoquean languages are agglutinative, as exemplified in (10), and they show extensive compounding, as in (11).

(10) Sierra Popoluca (Zoque)

| | | | |
|--------------------------------|---|--------------|--|
| <i>ʔit'iʔŋ o ʔjkaʔa ʔjnepa</i> | 'he _i is cutting with his _i instrument' | | |
| <i>ʔi</i> | <i>-tiŋ</i> | <i>-ʔoʔj</i> | <i>-kaʔ</i> <i>-ʔaʔj</i> <i>-ne</i> <i>-pa</i> |
| 3ERG | -cut | -ANTI | -INST -APPL -PROG -INCOMP |

(11) Coatlán Mixe

| | |
|------------------|--------------|
| <i>tsahptihk</i> | 'church' |
| <i>tsahp</i> | <i>-tihk</i> |
| sky | -house |

The family also shows widespread cliticization, for example, the person clitics in the Copainalá verb complex in (12a) and the stressed clitics of Sayuleño.

(12) (a) Copainalá Zoque

| | | | |
|---------------------|-------------------|-----------|-----------------------|
| <i>mbjojumih</i> | 'you ran' | | |
| <i>nj</i> | <i>-poj</i> | <i>-u</i> | = <i>mih</i> |
| 2ABS | -run | -COMP | = 2ABS |
| <i>nimih mbjoju</i> | 'you are running' | | |
| <i>ni</i> | = <i>mih</i> | <i>nj</i> | <i>-poj</i> <i>-u</i> |
| PROG | = 2ABS | 2ABS | -run -DEP |

(b) Sayuleño (Mixe)

| | | | |
|-----------------------|-------------|-------------|--------------------|
| <i>ʔik'naʔ</i> | 'the house' | | |
| <i>ʔik</i> | = | <i>'naʔ</i> | |
| house = DEF_(NON-FEM) | | | |
| <i>ʔif'pej</i> | | | 'he is going, too' |
| <i>ʔif</i> | <i>-p</i> | = | <i>'ej</i> |
| go | -INCOMP | = | also |

Mixe-Zoquean languages have two inflectional classes, nominal and verbal. Adjectivals are inflected like nominals.

Nominals inflect for number and possessor, and in southern Zoque, for case. The cases are absolutive and ergative/genitive.

- (13) (a) number
ʔihkat 'houses' Sayuleño (Mixe)
ʔik-bat
house-PL
(b) possessor
ŋguj 'my stick' Texistepec (Zoque)
n *-kuj*
1POSS -stick/tree
(c) case
pin 'man (*abs*)' Copainalá Zoque
piʔnis 'man (*erg/poss*)'

Nominals also have derivational forms with adpositional meanings, many built on reflexes of the protomorpheme **-mi* 'locative.' There are more adpositionals in southern Zoque and fewer in Mixe.

(14) Francisco León Zoque

| | |
|--------------------|------------------|
| <i>kumguʔjomo</i> | 'in/to the town' |
| <i>kumkuj-ʔomo</i> | |
| town-LOC | |

Verbs inflect for aspect, subordination, and person and number of both subject and object.

The aspects are incomplete and complete with allomorphs registering subordination. Most languages also have forms for future. In Zoque, the subordinate forms are only used in auxiliary constructions.

(15) (a) Oluta (Mixe)

| | independent | subordinate | |
|------------|------------------|--------------------|---------------|
| incomplete | <i>ti'kajp</i> | <i>tin'kaje</i> | 'I am eating' |
| complete | <i>ti'kaju</i> | <i>tin'kaji</i> | 'I ate' |
| future | <i>ti'ka'jam</i> | <i>tin'ka'jaʔn</i> | 'I will eat' |

(b) Copainalá Zoque

| | |
|---------------------------------|---------------------|
| independent | subordinate |
| <i>pojpaʔih</i> | <i>baʔnih mboje</i> |
| 'I'm running'/'I'm not running' | |
| <i>pojuʔih</i> | <i>hapjojaʔih</i> |
| 'I ran'/'I didn't run' | |
| <i>maŋbaʔih poju</i> | – |
| 'I will run' | |

The persons are first, second, and third, with a contrast of inclusive and exclusive in the first person plural.

(16) Sierra Popoluca (Zoque)

| | singular | dual |
|----------------|---------------------|-------------------------------|
| 'I go' | <i>ʔanikpa</i> | |
| 'you (sg.) go' | <i>mipikpa</i> | 'you and I go' <i>tanikpa</i> |
| 'he/she goes' | <i>nikpa</i> | |
| | plural | |
| 'we (ex.) go' | <i>ʔanik taʔmpa</i> | |
| 'we (in.) go' | <i>tanik taʔmpa</i> | |
| 'you (pl.) go' | <i>mipik taʔmpa</i> | |
| 'they go' | <i>nikyahpa</i> | |

Transitive verbs are marked inflected for both subject and object (or ergative and absolutive).

(17) Sierra Populca

| | | | |
|--------------|-----------------|------------------|------------------|
| | ‘... me’ | ‘... you’ | ‘... him’ |
| ‘I love :’ | – | <i>mipt’ojpa</i> | <i>?antojpa</i> |
| ‘you love :’ | <i>?antojpa</i> | – | <i>?iñt’ojpa</i> |
| ‘he loves :’ | <i>?atojpa</i> | <i>mit’ojpa</i> | <i>?it’ojpa</i> |

In the Mixean branch there is inersive person marking, including some true inverse systems, as in Sayuleño. Such systems contain two distinct transitive forms for third person acts on third person, a direct form in which the subject is in focus and an inverse form in which the object is in focus. The Sayuleño system is based on a person hierarchy that ranks first person above second person above third person. The inverse marker in Sayuleño is *-f-*, which has the third person allomorph *-gi-*.

(18) Sayuleño (Mixe)

| | | | |
|---------------|----------------|----------------|---|
| | ‘... me’ | ‘... you’ | ‘... him’ |
| ‘I hit ...’ | – | <i>tĩ’mojp</i> | <i>tiñ’mojp</i> |
| ‘you hit ...’ | <i>?ifmojp</i> | – | <i>?in’mojp</i> |
| ‘he hits ...’ | <i>tifmojp</i> | <i>?ifmojp</i> | <i>?in’mojp</i> (direct)/ <i>?igi’mojp</i> (inverse) |

As in some other Meso-American language families, there is a special class of verbs that are used to refer to the positions of people and objects. These are built on the reflexes of the derivational suffix **-nay-*. In most cases, the lexical roots found in this construction cannot be used in other combinations. The positions referred to are frequently semantically complex.

(19) Sayuleño (Mixe)

| | | |
|---------------------|-------------------------------|-----------|
| <i>'wehnap</i> | ‘he sits with his legs apart’ | |
| <i>'weh</i> | <i>-naY</i> | <i>-p</i> |
| sit with legs apart | -POSITIONAL | -INCOMP |

Syntax

The syntax of the Mixe-Zoquean family is typologically typical of Meso-American. Most of the languages are headmarking ergative, with VSO word order.

(20) Copainalá

| | | |
|---|-------------------------|-------------------------|
| <i>jahki?mu</i> | <i>jomo?s</i> | <i>te? sapane</i> |
| <i>j</i> – <i>jah</i> – <i>ki?m</i> – <i>u</i> | <i>jomo</i> – <i>?s</i> | <i>te? sapane</i> |
| 3ERG – CAUS – go_up – | woman-ERG | DEFbanana |
| COMP | | |
| V | S | O |
| <i>paktihkohmo</i> | <i>wa?y</i> | <i>kjiñu</i> |
| <i>paktihk</i> – <i>ohmo</i> | <i>wa?y</i> <i>j</i> | – <i>kiñ</i> – <i>u</i> |
| attic- LOC | that | 3ABS – ripen – DEP |
| adv | | |
| ‘The woman took the bananas to the attic to ripen.’ | | |

The syntax of dependent verb forms differs between the two branches of the family. The Zoquean languages use the dependent inflections only in construction with auxiliaries. The Mixean languages

have a wider variety of constructions triggering dependent inflection, for example, in clauses with fronted adverbials.

(21) Totontepec Mixe

| | | |
|---------------------|--------------|----------------|
| <i>yam</i> | <i>?əts</i> | <i>mpə:ʔkf</i> |
| <i>'yam</i> | <i>'?əts</i> | <i>n</i> |
| here | 1 | 1ABS_DEP |
| ‘I’m resting here.’ | | |
| <i>-pə:ʔkf</i> | <i>-Ø</i> | |
| -rest | -INCOMP_DEP | |

In auxiliary constructions, the auxiliary has third singular subject and bears the aspect. The lexical verb bears the subject person/number marking and incomplete dependent aspect.

(22) Sayuleño

| | | |
|-------------------------------|---------------|-------------|
| <i>kifp</i> | <i>naʃ</i> | <i>kaɟa</i> |
| <i>Ø</i> | <i>-kif</i> | <i>-p</i> |
| 3ABS-finish-INCOMP | 1INCL_ERG_DEP | |
| ‘We (incl.) finished eating.’ | | |
| <i>-kaɟ-ka</i> | <i>Ø</i> | |
| -eat-PL-INCOMP_DEP | | |

Object incorporation is widespread throughout the family.

Distinguishing the Branches

Phonological development distinguishes the two branches of the family. Zoquean underwent two significant developments. First, the length contrasts were leveled.

| Mixean | Zoquean |
|---------------|--------------|
| Sayuleño | Totontepec |
| <i>mo:bk</i> | <i>mɔ:bk</i> |
| <i>tɪ:hts</i> | <i>tɔ:ts</i> |
| <i>mok</i> | <i>bok</i> |
| <i>tits</i> | <i>tits</i> |
| | ‘maize’ |
| | ‘tooth’ |

The contrastive length in Gulf Zoquean is a secondary development. The first syllable of disyllabic roots is lengthened if it is underlyingly open, e.g., Sierra Popoluca *ka:ma*, Copainalá *kama* ‘cornfield,’ proto-Mixe-Zoquean **kama*. Second, a class of proto-Mixe-Zoquean CV:ʔC roots show glottal allomorphy in Mixe, which was leveled to CVC roots in Zoquean.

| Mixean | Zoquean |
|--------------|---------------|
| Sayuleño | Coatlán |
| <i>?u:kp</i> | <i>?u:kp</i> |
| <i>?u?k</i> | <i>?u?uk</i> |
| <i>?ukpa</i> | <i>?ukpa</i> |
| <i>?uki</i> | <i>?u:ki?</i> |
| | ‘he drinks’ |
| | ‘drink imp’ |

There are a few regular morphological features that distinguish Mixean from Zoque, for example, the future. Throughout the Mixean branch, the future is formed on the reflexes of a verbal compound with the stem *wa:ʔn* ‘want’. In Gulf Zoque the future is one

of the meanings of the incompletive forms (-*pa*). In Southern Zoque, futures are auxiliary constructions.

- (25) Mixean Zoquean
Sayuleño Coatlán Copainalá Sierra Popoluca
ʔuʔkam ʔoʔokop maŋbaʔuhku 'he will drink' *ʔukpa*

A number of common words or usages also distinguish the two branches.

- (26) Mixean Zoquean
Sayuleño Coatlán Copainalá Sierra Popoluca
ʃihw ʃi: hama ha:ma 'sun, day'
ʔiʔwamp jwa:mb ʃump ʔiʃump 'he wants it'
'toʔaj 'toʔʃhaj jomo jo:mo 'woman'

Bibliography

- Campbell L R & Kaufman T (1976). 'A linguistic look at the Olmecs.' *American Antiquity* 41, 80–89.
- Clark L E (1981). *Diccionario popoluca de Oluta. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 25. México, DF: Instituto Lingüístico de Verano.
- Clark L E & de Clark N D (1995). *Diccionario popoluca de Sayula. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 4 (2nd ed.). México, DF: Instituto Lingüístico de Verano.
- Clark L E & Nordell N (1990). *Vocabulario español-popoluca de Texistepec*. Ms.
- Elson B F & Gutiérrez G D (1999). *Diccionario popoluca de la Sierra Veracruz. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 41. México, DF: Instituto Lingüístico de Verano.
- Engel R, de Engel M A & Alvarez J M (1987). *Diccionario zoque de Francisco León. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 30. México, DF: Instituto Lingüístico de Verano.
- Grimes B (ed.) (2004). *Ethnologue: languages of the world* (14th ed.). Dallas: Summer Institute of Linguistics.
- Harrison R, Harrison M & García H C (1981). *Diccionario zoque de Copainalá. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 23. México, DF: Instituto Lingüístico de Verano.
- Harrison R, Harrison M, Juárez F L & Ordoñez C (1984). *Diccionario zoque de Rayón. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 28. México, DF: Instituto Lingüístico de Verano.
- Hoogshagen S & de Hoogshagen H H (1993). *Diccionario mixe de Coatlán, Oaxaca. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 32. México, DF: Instituto Lingüístico de Verano.
- Kaufman T & Justeson J (2003). 'Epi-Olmec.' In Woodward R (ed.) *Encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press. 1071–1111.
- Schoenhals A & Schoenhals L C (1965). *Vocabulario mixe de Totontepec. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'* 14. México, DF: Instituto Lingüístico de Verano.
- Suárez J A (1983). *The Mesoamerican Indian languages*. Cambridge Language Surveys. Cambridge: Cambridge University Press.

Mobilian Jargon

E Drechsel, University of Hawai'i at Mānoa, Honolulu, HI, USA

© 2006 Elsevier Ltd. All rights reserved.

Mobilian Jargon or the Chickasaw-Choctaw trade language was a Muskogean-based pidgin of the lower Mississippi River valley, typologically comparable (although unrelated) to Chinook Jargon of north-western North America. Not to be confused with Mobilian proper of southern Alabama, whose genetic classification has remained in doubt until today, Mobilian *Jargon* (MJ) was a structurally and functionally reduced contact medium that drew principally on Muskogean as its source languages. MJ displayed a characteristic Muskogean phonology, including the lateral fricative /ɬ/ as in /ɬaʎo/ 'fish', even if the pronunciation of particular sounds ranged more widely than in Muskogean languages because of

second-language interferences, and because it lacked most of their morphophonological complexities (*see Muskogean Languages*). Whereas early grammatical descriptions have presented MJ as a reduced form of Choctaw (Western Muskogean), it actually has revealed considerable lexical variation over space and time, reflecting the influences of its speakers' first languages. Among the primary other sources were Eastern Muskogean languages such as Alabama and Koasati and quite possibly Muskogee. Several 'exotic' loans from northern Algonquian languages with only a few from European sources suggest that MJ's range of variation extended farther north than the available linguistic evidence indicates, while remaining comparatively immune to European influences. A closer examination of MJ syntax demonstrates that whereas some constructions could derive from Choctaw or Chickasaw (Western Muskogean) equivalents, others displayed fundamental syntactic differences in need of their

own explanations. (The initial dagger indicates that these MJ sentences were not actual recordings, but corresponding word-for-word *reconstitutions* of such to match the Choctaw constructions, illustrating both the similarities and differences between the pidgin and Western Muskogean.)

- (1) MJ: †hattak išno pisa taha
 man you see PAST
 Choctaw: hattak iš- pisa tuk
 man 2 sing ACT see PAST
 (NOM)
 (Jacob *et al.*, 1977: 65)
 ‘You saw a man.’
- (2) MJ: †išno iti ino čali taha
 you wood I cut PAST
 Choctaw: iti chī- chāli -li -tok
 wood 2 sing cut 1 sing PAST
 DAT
 (Davies, 1986: 41)
 ‘I cut wood for you.’
- MJ: †ofi ino banna ino yimmi
 dog I want I believe
 Ch: ofi sa- banna yimmi -li -h
 dog 1 sing want believe 1 sing PRED
 ACC NOM
 (PAT) (ACT)
 (Davies, 1986: 71)
 ‘I believe I want a dog.’

Without some means of formally marking case in words, MJ speakers could rely only on word order to identify the grammatical functions of its sentence parts, which was X/OsV and quite possibly X/OSV (with no actual attestations available for two- or multiple-argument constructions, including nominal subjects). As the above examples illustrate, MJ’s unique word order probably derived from Muskogean constructions consisting of a noun plus a verb with a prefix, the latter of which speakers replaced with etymologically independent and full pronouns in their languages and which came to function as true grammatical subjects in MJ, although not necessarily as agents. MJ thus followed a grammatical pattern that was fundamentally Muskogean, even if structurally analytic and ultimately unintelligible to Muskogean speakers for larger constructions, unless they had considerable prior exposure to it.

Alternatively known as *anompa īla* (‘other/different/strange talk’), *yama* (‘yes, right, alright, indeed’), and *yoka anompa* (‘servant/slave talk’), MJ performed

not only as a true pidgin in multilingual contexts with non-Muskogean (as in intertribal gatherings, pantribal alliances, and intertribal and colonial trade), but also in bilingual contacts with distant peoples, including Algonquians, Siouans (such as the Osage), and eventually Europeans (as on long-distance travels, on diplomatic missions, in intertribal and interethnic marriages, and in the European employment of Indians). By the early 18th century, MJ moreover assumed meta-communicative functions of a sociolinguistic buffer against overly eager outsiders: its use helped confirm the native identity of its speakers, thus providing a safeguard against continuously threatening enslavement, while at the same time protecting their privacy against intrusions by missionaries, immigrant settlers, government officials, and anthropologists.

MJ’s structure and functions raises questions about its origin. Because the linguistically and culturally fairly uniform Muskogean had little reason to develop a pidgin, early analyses favored a colonial origin of MJ (Crawford, 1978). I have since proposed a pre-European origin for MJ on grounds of three interrelated arguments: its indigenous grammar and lexicon, its well-established use in diverse native interlingual contexts, and its geographic distribution closely overlapping with that of the linguistically diverse, but socioculturally quite uniform, paramount chiefdoms of the pre-Columbian moundbuilders known by archaeologists as the Mississippian Complex (Drechsel, 1996, 1997).

Bibliography

- Crawford J (1978). *The Mobilian trade language*. Knoxville: University of Tennessee Press.
- Davies W D (1986). *Choctaw verb agreement and universal grammar*. Dordrecht: D. Reidel Publishing Company.
- Drechsel E J (1996). ‘An integrated vocabulary of Mobilian Jargon, a Native American pidgin of the Mississippi valley.’ *Anthropological Linguistics* 38(2), 248–354.
- Drechsel E J (1997). *Mobilian Jargon. Linguistic and sociohistorical aspects of Native American pidgin* (Oxford Studies in Language Contact). Oxford: Clarendon Press.
- Jacob B, Nicklas D & Spencer B L (1977). *Introduction to Choctaw*. Durant, Oklahoma: Choctaw Bilingual Education Program, Southeastern Oklahoma State University.

Mon

J Watkins, School of Oriental and African Studies,
London, UK

© 2006 Elsevier Ltd. All rights reserved.

Mon is the principal language in the Monic sub-branch of Mon-Khmer languages, which form the bulk of the Austroasiatic language phylum. The near-extinct language Nyah Kur, spoken in Thailand, seems to be similar to Old Mon.

Mon is spoken in Burma/Myanmar and Thailand. In southeastern Burma, the Mon-speaking population lives in the area from Thaton across the lower Salween river area and down the coastal strip as far as Ye. Mon villages are interspersed with those of Burmans/Bamar and Karen/Kayin. The speakers of Mon in Thailand are thinly scattered in provinces surrounding Bangkok.

It is very difficult to put a precise figure on the number of people who speak Mon, not least because there is a large category of people who identify themselves as ethnically Mon but who do not speak the language. A further problem is the general lack of demographic data. Bauer (1990) undertakes a detailed analysis of the available information and concludes that there are probably one million Mon speakers, though this figure incorporates various degrees of bilingualism. Most of the Mon-speaking population is bilingual in Burmese or Thai. Of this million, roughly 50 000 – about 5% – reside in Thailand, and the remaining large majority resides in Burma. In Thailand, Mon have become heavily culturally and ethnically assimilated through extensive intermarriage between Mon and Thai, although being of Mon descent seems to carry some prestige.

Many sources suggest that the use of Mon is in decline in Thailand and possibly also in Burma, though this may not in fact be the case; again, the information available is confusing and contradictory. It appears that Mon is not yet a dying language. Although the government in Burma may have restricted teaching and official use of Mon, there are many villages in the middle and lower parts of the Mon state in Burma in which Mon is the only language spoken by people who are mostly literate in Mon, and it is predicted that Mon will continue to be a significant language in the region. Literacy rates in Mon are skewed toward men, as literacy is supported by monastery teaching, which is only partly accessible to girls. Mon language education organized by the New Mon State Party in Burma may redress the balance.

Mon is the ethnonym by which the Mon refer to themselves. This name can be traced back to Khmer

texts from the Sixth through the early twelfth-century Mon inscriptions at Pagan in Burma. The name 'Mon' – which in Mon is မန် / မှ် / မောန် [mòn] – is derived from the Old Mon RMEÑ, attested in an inscription dating from 1102 C.E. The initial RM-simplified to M- around the sixteenth century. The other names by which the Mon are known are the Burmese term *Talaing* တလိၼ်, considered derogatory, and *Peguan*, a geographical–historical name derived from Pegu, the ancient Mon capital.

The records of Mon cover a period from the sixth century to the present day. Certain collections show the state of the language at various times. The best-represented historical periods are the eleventh and twelfth and the fifteenth centuries C.E. Modern Mon dates from the mid-eighteenth century onward. Mon has influenced the major languages of mainland Southeast Asia, in particular Karen, Burmese, and Thai, all of which borrowed words from Mon in the first half of the second millennium C.E.

Modern spoken Mon differs considerably from written, literary Mon in phonology, lexicon, and syntax, to the extent that it is possible to consider literary Mon a source of loanwords in spoken language, just as much as Pali and Sanskrit. Early descriptions of Mon tended to ignore the difference between the colloquial and literary forms of the language. The two appear to have diverged from the sixteenth century, and both forms of the language have developed independent inconsistencies but remain inextricably intertwined. This complex situation has resulted in a high degree of ambiguity and redundancy in the orthography. For instance, the initial consonant of the first syllable in a typical disyllabic word is frequently h- in spoken Mon, though this pattern does not occur in literary Mon, which, despite being archaic, is the basis of the written language. The initial h- may be etymologically derived from a large number of different consonants, and so historically inaccurate variant spellings are common. For instance, the word pronounced *hədoh* 'strain, filter, sift' may be written in nine different ways: ခဉ်ဟ် KHADUIH, ဂဉ်ဟ် GADUIH, ထဉ်ဟ် THADUIH, ထဉ်ဟ် THADUIH, ထဉ်ဟ် THADUIH, ဒဉ်ဟ် DADUIH, ဖဉ်ဟ် PHADUIH, သဉ်ဟ် SADUIH, and ခဉ်ဟ် SDUIH. Literary Mon is an artificial construct that does not make sense without reference to spoken patterns. The mixture of the two forms has been described by the Mon scholar H. L. Shorto (1962: xvi) as 'a confusing scatter of seemingly aimless variation.'

Colloquial Mon exists in a range of dialects in Thailand and Burma, although all are mutually intelligible. The variety of Mon described here is that of

Shorto's *A dictionary of modern spoken Mon* (1962), which corresponds mostly with the Mon Rao of Nai Pan Hla's *An introduction to the Mon language* (1989), spoken by Mon in Burma.

There are a number of Mon language sites on the Internet. Some are associated with Mon political organizations, and some with news and information, such as ကောဝ် /kaowao/ *Kao Wao* ('black cuckoo') and သွင်တိုင် /hnoŋ taŋ/ *Guiding Star*. At the time of writing, Mon language is not included in Unicode or other international standards, and so, Mon text is displayed on the Internet as graphics.

Like the scripts used to write Burmese, Shan, and certain other languages of Burma/Myanmar, Mon is written with a script derived from the Indic scripts that spread with Buddhism to continental Southeast Asia in the early part of the first millennium C.E. The Mon were the dominant group in the south of an area that now straddles the border between Thailand and Burma/Myanmar and were centered in two areas, one along the Chaopraya River in Thailand, and the other in the Moulmein–Pegu (Mawlamyine–Bago) area of lower Burma, with a principle center at Thaton on the east coast of the Gulf of Martaban.

The oldest Mon inscription, found at Lopburi in Thailand, dates from the eighth century and is written

in the Southern Indian Pallava script. Because there are inscriptions in Pyu, an important language of the period, which predate those in Mon, it may be that the Mon borrowed the writing system from the Pyu. However, given, first, that Mon script bears a closer resemblance to Pallava script, and second, that the Mon city Thaton was itself a major Buddhist center, it may be that the Mon borrowed the writing system directly from India independent of the Pyu. It is thought that Mon scribes brought to the city of Pagan after the Mon were defeated by the Burmese king Anawratha in 1057 C.E. resulted in Mon script being adapted to write Burmese, though this theory has been disputed in recent research.

The Mon writing system is essentially the same as the Burmese one in appearance, but with certain modifications and some elements unique to Mon. The system is best suited to writing the Indic languages for which its parent scripts were first designed, such as Pali and, with a few extra symbols not shown here, Sanskrit.

Mon is not a tonal language like its geographical neighbors, but vowels in Mon occur mainly in pairs distinguished only by a quasitonal distinction known as a register distinction, which is found in many Mon-Khmer languages, including Wa and Cambodian (Khmer). First or 'head' register is characterized by clear voice quality and is associated with more peripheral vowels; second or 'chest' register is characterized by breathy voice (transcribed here with a grave accent, `), a general laxness of the speech organs, and centralization of the vowels.

The consonants of the Mon writing system are set out in Table 1. As in Cambodian (see Henderson 1952), each consonant is associated with one of the two registers, and the reading of a vowel in a given context is determined by the register designation of the consonant that governs it – usually the initial consonant of a syllable. The minimal pair in Table 2 illustrates the contrast.

Mon, like many Mon-Khmer languages, has a rich array of vowels. The word-initial and word-internal vowel symbols of Mon are shown and transliterated in Table 3. Table 4 shows how each of these is pronounced in head and chest register.

One of the distinctive features of Mon-Khmer phonology is the prevalence of relatively unrestricted initial consonant clusters. Mon is no exception, though

Table 1 The consonants of Mon transliterated and transcribed; consonants associated with second ('chest') register are bold

| Mon script | Transliteration | | | | | Transcription | | | | |
|------------|-----------------|---|---|-----|--|---------------|-----------------|----|-----------------|----|
| က | ခ | ဂ | ဃ | င | | k | k ^h | k | k ^h | ŋ |
| စ | ဆ | ဇ | ဈ | ည | | tɕ | tɕ ^h | tɕ | tɕ ^h | ɲ |
| တ | ထ | ဒ | ဓ | န | | t | t ^h | d | t ^h | n |
| တ | ထ | ဒ | ဓ | န | | t | t ^h | t | t ^h | n |
| ပ | ဖ | ဗ | ဘ | မ | | p | p ^h | p | p ^h | m |
| ယ | ရ | လ | ဝ | သ | | y | r | l | w | s |
| ဟ | ဠ | ဣ | အ | ဗ္ဗ | | h | l | ɔ̃ | ʔ | ɔ̃ |

Table 2 A minimal pair illustrating the register contrast in Mon

| | Mon script | Transliteration | Transcription | Phonetic detail | Gloss |
|----------------|------------|-----------------|---------------|-----------------------|-------|
| Head register | ကျင် | KLUN | /klɔ̃ŋ/ | [klɔ̃ ^h ŋ] | come |
| Chest register | ဂျင် | GLUN | /klɔ̃ŋ/ | [klɔ̃ ^h ŋ] | boat |

Table 3 Mon word-initial and word-internal vowel symbols

| Word-initial | | Word-internal | | Transliteration | |
|--------------|-----|---------------|---|-----------------|----|
| အ | အာ | — | — | A | Ā |
| ဣ | ဣိ | ◌ | ◌ | I | Ī |
| ဥ | ဥါ | — | — | U | Ū |
| ဧ | အဲ | ◌ | ◌ | E | AI |
| ဩ | အံ | ◌ | ◌ | O | AÓ |
| အံ | အံး | ◌ | ◌ | AM | AH |

Table 4 Mon vowels in head and chest register

| First ('head') register | | | Second ('chest') register | | |
|-------------------------|-----------|-------|---------------------------|-----------|-------|
| Mon | Translit. | Pron. | Mon | Translit. | Pron. |
| က | KA | ka? | ဂ | GA | kè? |
| ကာ | KĀ | ka | ဂါ | GĀ | kèa |
| ကိ | KI | kəe? | ဂီ | GI | kí? |
| ကီ | KĪ | kəe | ဂိ | GĪ | kì |
| ကု | KU | kao? | ဂူ | GU | kù? |
| ကူ | KŪ | kao | ဂု | GŪ | kú |
| ကေ | KE | ke | ဂေ | GE | kè |
| ကဲ | KAI | koa | ဂဲ | GAI | kòa |
| ကာ | KO | kao | ဂါ | GO | kà |
| ကော | KAO | kao | ဂေါ | GAO | kèa |
| ကံ | KAÓ | kòm | ဂံ | GAÓ | kòm |
| ကံး | KAH | kah | ဂံး | GAH | kèh |

Translit. = transliteration; Pron. = pronunciation.

as mentioned above, in the colloquial language many such clusters are not pronounced as spelt. Mon script writes the second element of such clusters with a subscript form of the consonant, as shown in **Table 5**.

In addition to this already rich inventory of open syllable vowels, Mon features a large array of additional syllable rhymes that involve further diphthongs and final consonants (see Nai Pan Hla [1989]) for further details.

The following example sentences in Mon illustrate some of the basic properties of Mon syntax. Modifiers generally follow what they modify (1); SVO (subject–verb–object) order is observed (2); subject pronouns are dropped (3). Unusually for the languages of mainland Southeast Asia, Mon makes only limited use of classifiers and has a system of plural marking.

Table 5 Subscript forms of Mon consonants

| | | | | | |
|----|----|----|------|-------|--------------------------|
| င | အါ | တူ | TNA | təŋa? | crossbow |
| ည | အါ | ည | KNA | ŋa? | (honorific prefix) |
| ချ | အါ | ဆ | CDA | da? | span |
| န | အါ | ပ | PNA | pəna? | pretence |
| မ | အါ | တူ | KMA | kəma? | insect |
| ယ | အါ | ကျ | KYA | ca? | be defeated |
| ရ | အါ | ဂြ | GRA | krè? | gather up |
| လ | အါ | တူ | KLA | kla? | tiger |
| ဝ | အါ | တူ | KWA | kwa? | short of stature |
| စ | အါ | ထွ | THBA | həba? | lift up |
| ဟ | အါ | လှ | LHA | hla? | leaf (also spelt လှ SLA) |

*Note the divergence of spoken forms from written spellings in these words.

(1) ရဲ အဲ
RAI AI
ròa ?òa
friend I
'My friend'

(2) အဲ ရာန် ယာတ် မွဲ အုပ်
AI RAN YAT MVAI 'UP
?òa ràn yàt mòa ?up
I buy cloth one length
'I bought a length of cloth.'

(3) မွဲင် ပဲ ချင် လှိုင်
DMĀN PDAI DUN LGUN
mòŋ dɔa dɔŋ təkɔŋ
live in town Rangoon
'(He) lives in Rangoon/Yangon.'

Bibliography

Bauer C (1982). *Morphology and syntax of spoken Mon*, Ph.D. thesis, University of London (SOAS).

- Bauer C (1986). 'The verb in spoken Mon.' *Mon-Khmer Studies* 15, 87–110.
- Bauer C (1990). 'Language and Ethnicity: the Mon in Burma and Thailand.' In Wijeyewardene G (ed.) *Ethnic groups across national boundaries in mainland Southeast Asia*. Singapore: Institute of Southeast Asian Studies. 14–47.
- Diffloth G (1984). *The Dvaravati Old Mon language and Nyah Kur*. Monic Language Studies I. Bangkok: Chulalongkorn University.
- Diffloth G (1985). 'The registers of Mon vs. the spectrographist's tones.' *UCLA Working Papers in Phonetics* 60, 55–58.
- Guillon E (1976). 'Some aspects of Mon syntax.' In Jenner P, Thompson L & Starosta S (eds.) *Austroasiatic studies*. Oceanic linguistics special publication 13, vol. 1. Honolulu: University of Hawaii Press. 407–422.
- Guillon E (1999). *The Mons: a civilisation of Southeast Asia*. Bangkok: The Siam Society.
- Halliday R (1922). *A Mon-English dictionary*. Bangkok: Siam Society.
- Haswell J M (1901). *Grammatical notes and vocabulary of the Peguan language*. Rangoon: American Baptist Mission.
- Henderson E J A (1952). 'The main features of Cambodian pronunciation.' *Bulletin of the School of Oriental and African Studies* 14, 149–174.
- Huffman F (1990). 'Burmese Mon, Thai Mon, and Nyah Kur: a synchronic comparison.' *Mon-Khmer Studies* 16–17, 31–84.
- Luang Thongkum T (1987). 'Another look at the register distinction in Mon.' *UCLA Working Papers in Phonetics* 67, 132–165.
- Pan Hla N (1989). *An introduction to Mon language*. Kyoto: Kyoto University Center for Southeast Asian Studies.
- Shorto H L (1962). *A dictionary of modern spoken Mon*. London: Oxford University Press.
- Shorto H L (1966). 'Mon vowel systems: a problem in phonological statement.' In Bazell C, Catford J, Halliday M & Robins R (eds.) *In memory of J. R. Firth*. London: Longmans. 398–409.
- Shorto H L (1971). *A dictionary of the Mon inscriptions from the sixth to the sixteenth centuries*. London: Oxford University Press.
- Smalley W (1994). *Linguistic diversity and national unity: language ecology in Thailand*. Chicago: University of Chicago Press.
- South A (2003). *Mon nationalism and civil war in Burma: the golden sheldrake*. London: RoutledgeCurzon.
- Tun Way N (1997). *The modern English–Mon dictionary*. Bangkok: Open Society Institute Burma Project.

Mongolic Languages

J A Janhunen, University of Helsinki, Helsinki, Finland

© 2006 Elsevier Ltd. All rights reserved.

Definition

Mongolic is the technical term for the group of languages (conventionally also known as Mongolian) spoken by the linguistic descendants of the historical Mongols, a medieval ethnic group that created the political entity known as the Mongol Empire (early 13th to late 14th centuries). The historical Mongols were pastoral nomads who started their expansion from a relatively compact homeland centered in northeastern Mongolia and northwestern Manchuria, but under the Mongol Empire they were dispersed all over central and northeastern Asia, where populations speaking Mongolic languages still survive today. By the degree of dispersal and diversification, Mongolic may be characterized as a medium-large language family. The principal neighbors of Mongolic in premodern times have included Turkic (in the west), Tungusic (in the northeast), (Mandarin) Chinese (in the southeast), and Tibetan (in the south). There is

also both direct and indirect information of a group of historical and protohistorical languages, termed Para-Mongolic, which were collaterally related to the language of the historical Mongols. Due to problems of deciphering and linguistic analysis, these languages have not yet been incorporated into the corpus of Mongolic comparative studies.

Distribution

The geographical core area of the Mongolic family coincides with the historical and geopolitical entity of Mongolia, covering both Outer Mongolia (the modern independent state of Mongolia) and Inner Mongolia (an autonomous region within China). In terms of physical geography, this area comprises the steppes and adjoining forested highlands of the Mongolian Plateau, the Ordos region south of the Yellow River, and the Gobi Desert. In the north the Mongolic area extends to the Baikal region in the Siberian forest zone (Russia), in the east to the plains of the Manchurian provinces (China), in the south to the Gansu corridor and the Amdo (Kokonor) region (China/Tibet), and in the west to the Jungarian

basin in eastern Turkestan or Xinjiang (China). Separate areas and relicts of Mongolic-speaking populations are found in Afghanistan and the Caspian region (Russia), as well as in some parts of middle Asia (Kazakhstan and Kyrgyzstan). The area of the historical Para-Mongolic languages was centered on southwestern Manchuria.

Time Depth

Mongolic (excluding Para-Mongolic) is a well-delimited family of a dozen closely related languages, which derive from a relatively shallow and dialectally coherent protolanguage, termed Proto-Mongolic. In view of the circumstances underlying the dispersal of the Mongolic languages, the diachronic depth of Proto-Mongolic must be less than 1000 years. Although Proto-Mongolic is, by definition, a hypothetical construction that can only be approached by the method of comparative linguistics, it must have been close to the language of the historical Mongols, also known as Middle Mongol, which is actually documented in a variety of sources contemporary with the Mongol Empire. To some extent, these sources illustrate the gradual dialectal diversification of Middle Mongol, which ultimately led to the separation into the modern Mongolic languages. Contacts between some of the individual Mongolic languages have continued until modern times, making their boundaries with one another, in some cases, fuzzy.

Genetic Status

Mongolic is often classified as a branch of the Altaic language family, which is also supposed to include Turkic and Tungusic, as well as Korean(ic) and Japonic (Japanese-Ryukyu). Although the conception of an Altaic genetic unity still has adherents, modern research has demonstrated that the relationships of Mongolic with the other languages of the Altaic complex are best explained in terms of a complex and multilayered network of historical and pre-historical areal contacts. Most important, Mongolic has over several millennia been in contact with Turkic (in the west) and Tungusic (in the east), resulting in a considerable corpus of shared structural properties and linguistic substance among all the three language families. This interaction has continued up to the present day in some regions (Siberia, Manchuria, Eastern Turkestan, and Amdo). On a higher level, Mongolic also belongs to the areal and typological context of Ural-Altaic, which in addition comprises the Uralic language family. Both Altaic and Ural-Altaic remain relevant (and still insufficiently understood) concepts of areal linguistics and typology,

but in the genetic sense these terms may today be regarded as obsolete.

Classification

Due to the shallowness of Proto-Mongolic, the Mongolic languages are difficult to classify in terms of a clear-cut (binary) family tree. It is, however, possible to establish four relatively distinct branches of Mongolic (**Table 1**). Two of these branches comprise only one marginal language each: Dagur (Daur; in Manchuria) and Moghol (Moghli; in Afghanistan), which also seem to be the two Mongolic languages most distant from one another as determined by the number of shared isoglosses. The two other branches may be called Central Mongolic and Shirongolic. Central Mongolic covers a large coherent area centered on Mongolia, but it is differentiated into five distinct, although closely related, languages: Khamnigan Mongol (in the northeast), Buryat (Russia Buriat, Mongolia Buriat, and China Buriat; in the north), Mongol proper (in the center and east), Ordos (Peripheral Mongolian; in the south), and Oirat (Kalmyk-Oirat; in the west). Shirongolic also comprises a cluster of at least five distinct languages in the Amdo (Kokonor) region: Shira Yughur (also known as East Yughur), (Huzhu) Mongghul (Tu), (Minhe) Mangghuer (Tu), Bonan (Baoan), and Santa (Dongxiang). All of the Mongolic languages comprise a number of dialects and subdialects, some of which could linguistically also be counted as separate languages. Internal diversification is particularly conspicuous in Mongol proper, Buryat, and Huzhu Mongghul.

Typology

Proto-Mongolic may be reconstructed as a rather consistently agglutinative language with a sentence structure and suffixal morphology of the Ural-Altaic type.

Table 1 Classification of the Mongolic languages

| Branches | | Languages | Location |
|---------------------|------|----------------------|-------------|
| 1. Dagur(ic) | | (1) Dagur | Manchuria |
| 2. Central Mongolic | 2.1. | (2) Khamnigan Mongol | Manchuria |
| | 2.2. | (3) Buryat | Siberia |
| | 2.3. | (4) Mongol proper | Mongolia |
| | 2.4. | (5) Ordos | Ordos |
| | 2.5. | (6) Oirat | Jungaria |
| 3. Shirongolic | 3.1. | (7) Shira Yughur | Amdo |
| | 3.2. | (8) Huzhu Mongghul | Amdo |
| | 3.3. | (9) Minhe Mangghuer | Amdo |
| | 3.4. | (10) Bonan | Amdo |
| | 3.5. | (11) Santa | Amdo |
| 4. Moghol(ic) | | (12) Moghol | Afghanistan |

This conclusion is confirmed by actual information from Middle Mongol, and similar typology is still synchronically observed in both Dagur and the languages of the Central Mongolic branch. The two other branches of Mongolic have, however, undergone fundamental changes in their typological orientation. Moghol, under the influence of the local Iranian languages, has developed a large number of non-Mongolic features, including prepositions, conjunctions, and new inflexional categories. The Shirongolic languages, on the other hand, have been influenced by both local Chinese and local (Amdo) Tibetan in a complex framework of areal interaction that may be termed the Amdo Sprachbund (also known as the Qinghai-Gansu Sprachbund).

Literary Use

Since the time of the Mongol Empire, the principal literary language of Mongols has been Written Mongol (also known as Literary Mongol), written in a Semitic script adopted from the ancient Uighurs of eastern Turkestan. The early forms of Written Mongol were close to Middle Mongol, also recorded in other scripts, whereas the later forms become increasingly close to the spoken dialects of Mongol proper, the most important modern Mongolic language. An adaptation of Written Mongol known as Written Oirat was introduced as a written language of the Oirat speakers in 1648, but otherwise Written Mongol has been used by the speakers of all the Central Mongolic languages. Written Mongol was the official written language of Mongolia until the 1940s, and it still has an official position in Inner Mongolia. Outside of Inner Mongolia, it has, however, been replaced by new literary languages (in Cyrillic and Roman scripts) based on the local vernaculars.

Political Status

Written Mongol used to be one of the five official languages of the Manchu Empire of China (1644–1911), and it still remains one of the few minority languages in which public education is available in China.

The dominant spoken language of the Mongols in both Outer and Inner Mongolia is Mongol proper. The principal dialectal form of Mongol proper in Outer Mongolia is Khalkha (Halh Mongolian), which has been developed as a literary language (in Cyrillic script) since the 1940s and is today the official state language of Mongolia.

Within the Russian Federation (in Siberia), Buryat has an official status (including a literary standard in Cyrillic script) in the Republic of Buryatia and

other Buryat administrative areas, whereas a diaspora variety of Oirat known as Kalmuck (or Kalmyk) has a similar status in the Republic of Kalmykia (in the Caspian region). Other Mongolic languages have no official position, but experiments with modern literary languages (in Roman script) for Shira Yughur, Huzhu Mongghul, Minhe Mangghuer, and Santa are being made.

Demography

The total number of Mongolic speakers today is ca. 5–7 million. Most of these speak dialects of Mongol proper, of whom ca. 2 million live in Mongolia and ca. 3–4 million in Inner Mongolia (and other parts of China). Other relatively large Mongolic populations are those speaking Santa (ca. 600 000), Oirat (including Kalmuck, ca. 300 000), Buryat (ca. 300 000), Dagur (ca. 100 000), Ordos (ca. 100 000), and Huzhu Mongghul (ca. 100 000). The other languages are spoken by considerably smaller populations: Minhe Mangghuer (ca. 30 000), Bonan (ca. 10 000), Shira Yughur (ca. 3000), and Khamnigan Mongol (ca. 2000). Moghol is spoken by only a few individuals, if any.

With regard to linguistic vigorousness, there are considerable differences among the Mongolic languages. Whereas Moghol is moribund or nearly extinct, some groups of Khamnigan Mongol and Bonan are still viable in spite of their small numbers of speakers. Huzhu Mongghul is rapidly declining, whereas Minhe Mangghuer seems to be stable for the time being. The numbers of Buryat, Dagur, and Ordos speakers are diminishing due to assimilation by Chinese, Russian, and Mongol proper. Mongol proper is also rapidly losing ground in Inner Mongolia due to assimilation by Chinese. In this situation, Santa, a little investigated Mongolic language spoken by a compact and mainly monolingual Moslem population, with a low general level of literacy and education, is possibly the most vigorous and demographically the most rapidly growing Mongolic language today.

History of Research

Much of the early work on Mongolic focused on the philological analysis of Written Mongol texts. Isaac-Jacob Schmidt, working in service to Russia, was the first to publish a scientific grammar and dictionary of Written Mongol in the 1830s. Linguistic work on living Mongolic languages was initiated in the 1840s by the Finnish ethnolinguist M. A. Castrén (who worked on Buryat) and continued in the late 19th and early 20th centuries by G. J. Ramstedt (on Khalkha, Oirat/Kalmuck, and Moghol), A. D. Rudnev (on dialects of

Buryat and Mongol proper), Nicholas Poppe (on Buryat, Khalkha, and Dagur), and Antoine Mostaert (on Ordos). The documentation of Moghol was completed by Michael Weiers in the 1970s. The last major blank spot in Mongolic studies was the Shiron-golic branch. After the pioneering contributions by Antoine Mostaert (Huzhu Mongghul) in the 1920s and 1930s, the Shiron-golic languages were studied by a Sino-Russian expedition under the leadership of B. Kh. Todaeva in the 1950s and by an Inner Mongolian expedition in the 1980s. Even so, material on some Shiron-golic languages and dialects has been extremely scarce until the present day. The last major Mongolic language to be documented was Minhe Mangghuer, described by Keith W. Slater (2003). The most up-to-date general work, with grammatical sketches of all Mongolic languages, was edited by Juha Janhunen (2003).

In the field of diachrony, the focus was long on external comparisons in the Altaic framework. The main source on Mongolic comparative studies remains the work of Nicholas Poppe (1955), which, unfortunately, has already become obsolete in some respects, especially as far as the languages of the Shiron-golic branch are concerned.

Bibliography

- Janhunen J (ed.) (2003). *The Mongolic languages*. London and New York: Routledge.
- Poppe N (1955). 'Mémoires de la Société Finno-Ougrienne 110.' *Introduction to Mongolian comparative studies*.
- Slater K W (2003). *A grammar of Mangghuer: a Mongolic language of China's Qinghai-Gansu Sprachbund*. London and New York: Routledge.

Mon-Khmer Languages

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

The Mon-Khmer languages constitute a disparate group of languages belonging to the Austro-Asiatic phylum spoken in a large area across Southeast Asia. The term 'Mon-Khmer' has several interpretations. One sense in includes all non-Munda Austroasiatic languages (except Nihali if this is in fact to be considered Austroasiatic at all; for more see **Austroasiatic Languages**). Another sense of Mon-Khmer excludes Nicobarese, while a further interpretation excludes both the Nicobarese and the Aslian branches. This latter one is the understanding of the term 'Mon-Khmer' relevant for this article.

Within this narrow-scope interpretation of Mon-Khmer, the following subgroups can be reckoned: Bahnaric, Katuic, Khasi(c), Khmeric, Khmuic, Monic, Palaung-Wa, Pearic, and Viet-Muong. The internal relations of these subgroups have yet to be adequately determined to the satisfaction of specialists; as such, no *Stammbaum* of Mon-Khmer can be here offered.

One proposal (Grimes, 2000) on the internal subgrouping of Mon-Khmer within the broadest understanding enumerated above, recognizes a major cleavage between Aslian and all other Mon-Khmer languages; these latter are further subdivided into an Eastern Mon-Khmer subgroup consisting of Bahnaric,

Katuic, Khmer, Pearic, Monic, and Nicobarese and a Northern Mon-Khmer group that consists of Khmuic, Khasic, Palaung-Wa, and the isolate Mang of Vietnam. Viet-Muong is considered a separate branch coordinate with other Mon-Khmer subgroups. In addition, a number of relatively recently identified Mon-Khmer languages of China and Vietnam either appear to be isolate branches or remain unclassified, e.g., Palyu or Bugar. These are briefly discussed below. Lower-level taxonomic subgroupings have also been offered, e.g., Katuic-Bahnaric within the Eastern Mon-Khmer branch. Further research will refine and revise the classification and internal relations of the Mon-Khmer languages. However, given the controversial nature of all but the highest-order taxonomic subdivisions within Mon-Khmer, a conservative approach is offered here.

Within the context of individual subgroups of Mon-Khmer, comments are offered below on total numbers of speakers, etc. However, a few general overall comments on certain salient linguistic features can be made. Many Mon-Khmer languages exhibit unusual or noteworthy phonological features, such as the predilection to 'sesquisyllabic' (one-and-a-half-syllable) words that consist of a major/full syllable and a minor/reduced syllable. This takes the shape of reduced + full (or minor + major) and yields words with atypical clustering in initial position; examples can be found even in the names of several Mon-Khmer languages (or subgroups), e.g., Khmer, Khmu, Sre, Mnong,

Mrabri, etc., Vowel systems among Mon-Khmer languages are frequently highly developed, with elaborate systems of back unrounded vowels, centralized vowels, etc. often in combination with various phonation types or register phenomena. Such phonation types include creaky voice, breathy voice, etc. This combination of large core vowel systems and phonation types yields exploded inventories of syllable nuclei and/or vowel phonemes in various individual Mon-Khmer languages. These rank among the largest, if not the largest, such inventories in the languages of the world.

While the phonological systems of Mon-Khmer languages are highly developed, the languages are relatively impoverished morphologically and tend toward isolating word structure. However, the presence in most languages of lexicalized derivational elements, as well as productive or active systems in such languages as Bahnar, suggests that Proto-Mon-Khmer may have been more morphologically rich than most of its daughter languages. Indeed, a range of affixational processes may be used in individual languages within the Mon-Khmer language family, such as the following examples of derived nouns from Palaung (Milne, 1921: 74–75): *ra-pʌn-hwɔːi* [NOM₁-NOM₂-finish] ‘completion’; *pʌn-ra-iːr* [NOM₂-NOM₁-hate] ‘loathing, abhorrence’; *ra-ka-rɔt* [NOM₁-NEG-arrive] ‘the not arriving’; etc. Morphologically rich verbs are also to be found in individual Mon-Khmer languages, as is found in the following Khasi form:

- (1) *ya-pin-sam-θya?*
 RECIP/DIST-CAUS-INCLIN-sleepy
 ‘together make (others) feel sleepy’
 (Nagaraja, 1985: 27)

Similar morphologically developed verb forms may also be found in such Mon-Khmer languages as Katu and Bahnar. Indeed, it is possible to find cognate processes of affixation across several branches of the family, which may thus be reconstructed back to the Mon-Khmer protolanguage. Such is the case with the derivation of causative verbs. It appears that Proto-Mon-Khmer utilized a causative prefix when the verb stem was monosyllabic, but an infix when the stem was sesquisyllabic or longer. Such a pattern is found for example in Mon (Old Mon and modern Mon), the Katuic language Kuy, and Khmu. Note that the infix allomorph is realized as a (syllabic) nasal in Khmu and Kuy but as a schwa in Mon.

- (2) Khmu
p-háan ‘kill’ *p-rəh* ‘raise’ *k-m-sés* ‘drop’
 ←*háan* ‘die’ ← *rəh* ‘rise’ ←*k-sés* ‘fall’
 t-m-lüuy ‘hang (something)’
 ←*t-luy* ‘hang’
 (Svantesson, 1983: 104)

- (3) Spoken Mon: *p*-vs. -ə- (<*m/C__C)
hum daik → *p-hum daik* *klāŋ* →
 ‘have a bath’ ‘bathe’ ‘be numerous’
 hə-lāŋ
 ‘increase’

(Bauer, 1989: 90)

- (4) Old Mon
kcət → *kəcət*
 ‘die’ ‘kill’
 (Bauer, 1990: 149)

- (5) Kuy
kəcet → *kəmcet*
 ‘die’ ‘kill’
 (Bauer, 1990: 149)

This pattern is cognate with the system seen in Nicobarese and Proto-Munda as well, and as such appears to be a formation dating all the way back to the Proto-Austroasiatic level; see Anderson and Zide (2001) for more on this.

Syntactically, Mon-Khmer languages, like many languages of greater southeast Asia, possesses sequences of verbs commonly referred to as serial verb constructions (Schiller, 1990). Examples of a verbal sequence of this type includes the following from Khmer and Ravüa of the Palaung-Wa branch:

- (6) Khmer
təu jɔk kasæet mɔk
go take newspaper come
 ‘go get the newspaper’
 (Schiller, 1990: 40)

- (7) Ravüa
ti me ho taw lik me pin ke-en
take you go send letter you accompany to-here
 ‘go take the letter and come back’
 (Schiller, 1990: 58)

Mon-Khmer languages are generally spoken in remote, hilly/mountainous, and isolated enclaves spread across northern Thailand, Laos, Cambodia, southern China, Myanmar, northeastern India, and Vietnam. In many instances they are spoken by only a few hundred or few thousand speakers. However, Mon-Khmer languages are also the national majority language of both Cambodia (Khmer) and Vietnam (the heavily Sinicized Vietnamese).

Bahnaric is a large group of minority languages spoken in southern central Vietnam, southern Laos, and northwestern Cambodia. The total number of all Bahnaric language speakers is likely less than one million. There are three or four major subdivisions within Bahnaric. Bahnaric clearly has Northern, Western, and Southern branches, to which is added by some researchers a Central Bahnaric branch as well. The entire Southern subgroup is spoken in Vietnam, as are all but one of the Northern Bahnaric

languages (Talieng is spoken in Laos). South Bahnaric is separated from the other Bahnaric languages by a group of Mainland Austronesian languages. West Bahnaric languages, on the other hand, are not found in Vietnam at all, but rather are dispersed throughout various enclaves in Laos and Cambodia. The Central Bahnaric languages, which includes Bahnar proper, is a disparate group of five languages scattered across Vietnam, Laos, and Cambodia.

(8) Bahnaric subgroups

Northern Bahnaric

East: Cua, Kayong
Takua

West: Halang Doan
Jeh, Halang
Rengao
Sedang, Hre; Monom, Todrah
Unclassified: Talieng, Trieng

Unclassified: Katua

Southern Bahnaric

Sre-Mnong: Mnong: Eastern Mnong
Central Mnong,
Southern Mnong

Sre: Koho, Maa
Stieng-Chrau: Chrau
Stieng (Bulo Stieng)

Western Bahnaric

Bräo-Kravet: Bräo (Lave), Kravet, Kru'ng
(Kru'ng 2), Sou

Laven: Laven

Nyaheun: Nyaheun

Oi-The: Jeng, Oi (Oy), Sapuan, Sok, The

Central Bahnaric

Bahnar, Romam (Vietnam)

Alak (Laos)

Lamam, Tampuan (Cambodia)

Languages or dialects within these various headings include the following:

(9) Dialects/languages within various Bahnaric-speaking groups

Stieng: Budip, Budeh, Bulach, Bulo (a.k.a.

Ke-dieng, Se-dieng, and Rmang)

Southern Mnong: Nong/Diq, Prong (Prang)/

Rbut; perhaps also Rahong, Bu Sre

Central Mnong: Preh, Budong, Burung, Dih Bri,
Bunor, Biat

Eastern Mnong: Rlam (Ro'lo'm), Gar, Kuenh,
Dliê Ruc, Ndee

Koho: Maaq, Sre, Tôla, Nop, Köyon, Cil (Kou
N'ho), Tring, Nohang, Lat/Lac, Riong, Pru,
Laya, Röda, Co Don, Kalop

Bahnar: Alakong, Tolo, Bonom, Golar, Jolong,
Kontum, Rongao (Rengao), Kon KO De,
Krem, Roh, To Sung, Hodrung, Hroi, 144
M'nhar

Sedang: Central Sedang, Greater Sedang, Daksut,
Kon Hring, Kotua

Tampuon: Kroi, Lamam/Rmam

It is not possible to accurately gauge the exact number of speakers of Bahnaric languages, due in part to the extensive displacement many experienced during the Vietnam War. Many Bahnaric language speakers have been influenced by and/or shifted to national languages such as Vietnamese or Khmer or locally dominant languages such as Rhadé (Rade) or Cham (Austronesian). The following constitute rough estimates only:

(10) Estimated number of speakers of Bahnaric languages

Koho 100 000+ (including 23 000–30 000 Sre,
30 000–40 000 Maaq, 14 000 Cil, 3000 Lac,
6000 Nop, 14 000 Riong, very few each of
Laya, Co Don, Tôla)

Bahnar 85 000?

Hre 80 000?

Stieng 48 000?

Sedang 40 000?

Central Mnong 23 000?

Chrau 20 000?

Boloven (Laven) 18 000?

Cua 15 000?

Southern Mnong 12 000?

Eastern Mnong 12 000?

Halang 10 000?

Jeh 10 000?

Brau/Lavé 3000?

Koyong < 3000?

Nha Hôn 2500??

Todrah, Alak, Takua, Cheng, Sapuan, Oi, Souq,

Pragar, Kayong, Bout, Duan ??

Note that Parkin (1991) considers Central Bahnaric languages to be North Bahnaric.

Katuic languages are spoken in the region where Laos, Cambodia, and Vietnam meet. There are two subgroups within Katuic, conventionally called Eastern and Western Katuic. The total number of speakers of Katuic languages is approximately 200 000–300 000. The main languages include Katu, with 20 000–30 000 speakers, Bru, with possibly as many as 80 000 speakers, Sô (Tro), with 10 000 speakers, Pacoh, with 15 000 speakers, Ta'oih, with perhaps 10 000 speakers, Souei (Proom), with around 10 000 speakers, Kataang, with at least 10 000 speakers, Kaleung, who have mainly shifted to Lao but who number perhaps 40 000, and finally Kuy, with possibly as many as 150 000 speakers. Other than the Katu proper, the Pacoh (and the closely related Phuong), and the Khua, who live in Vietnam, and the Kuy and Western Bru of Thailand (and northern Cambodia),

Katuic-language speakers live mainly in Laos. Many are undergoing shift to Lao.

- (11) Katuic languages
 Eastern Katuic
 Kasseng
 Kataang
 Katu–Kantu
 Ngeq–Khlör–Alak 2
 Pacoh–Phuong
 Lower Ta’oih–Upper Ta’oih–Ir–Ong
 Tareng
 Western Katuic
 Kuy–Nyeu
 Eastern Bru–Western Bru–Khua–Leun–
 Mangkong–Sapoin–So Tri–Sô

Khasi[c] is the only branch of the Mon-Khmer language family spoken in mainland India. For more details see **Khasi**.

Khmeric consists of two languages: Central or Standard Khmer, the national language of Cambodia, and Northern Khmer, spoken mainly across the border in Thailand. For more on Khmer see **Khmer**. Khmer has been attested since the 7th century and appears in at least four historical stages: Pre-Angkorian, Old Khmer, Middle Khmer, and Modern Khmer. There may be as many as seven million Khmer speakers.

The Khmuic subgroup of Mon-Khmer consists of approximately a dozen languages scattered across Laos, Vietnam, and Thailand, with small enclaves in Myanmar and China as well. The Khmuic branch is further subdivided into the following subgroups:

- (12) Khmuic languages
 Khao
 Khao (Vietnam)
 Bit (Laos)
 Mal-Khmu
 Khmu
 Khmu (Laos, Thailand, Vietnam, China,
 [Myanmar])
 Khuen (Laos)
 O’du (Vietnam)
 Mal-Phrai
 Lua’ (Thailand)
 Mal (Laos)
 Phrai (Thailand)
 Pray (Pray 3) (Thailand)
 Mlabri (Thailand)
 Xinh Mul
 Khang (Vietnam)
 Pong (Phong Kniang) (Laos)
 Puoc (Vietnam)

Apart from these designations, which are standard in the Western linguistic tradition, Khmuic-speaking

peoples are known by a plethora of local variants (Parkin, 1991: 96). These include the folk-etymologized Kha Mu in Laos (*kha* is a general term for subjugated ‘hill tribes’ in Laotian), sometimes also Phou Theng or local designations such as Thay Hay or Hok in Laos and Rook in Thailand; in China they were formerly known as Chaman, now Kemu; in Vietnam, the Khmu are often referred to as Xa Cau. The total number of Khmuic speakers is moderately large, with Khmu proper the largest group, having between 350 000 and 500 000 total speakers in numerous local variants. Mal-Phrai, also known as T’in, has perhaps 20 000 speakers and is probably the next largest Khmuic-speaking group. Many languages are spoken by very small populations, e.g., Khang Ai of Tay Bac Province, Vietnam, which may have fewer than 1000 speakers; a similar number is estimated for Bit of northern Laos, while Mlabri (also known as Mrabri and Yumbri) may have as few as 200 speakers.

One group that deserves special mention here are the Lamet, who are sufficiently Khmuized linguistically and culturally to make their classification unclear. It is possible that they were originally speakers of a Palaungic language but their exact classificatory status remains open.

The Monic branch of Mon-Khmer consists of just two languages, Mon of Myanmar and Thailand and Nyahkur of Thailand. Mon, like Khmer, has a long literary tradition, with texts dating back 1000 years to the time when the Mon ruled an empire in this region; isolated inscriptional sources date back as far as the 7th century. Ethnic Mon may number nearly half a million, but the total number of speakers is significantly less, possibly only a tenth of that figure. The Nyahkur, on the other hand, total no more than a few thousand speakers, and probably represent the remnant of an old Mon kingdom of southern Thailand. In Thai they are called Chaubon; both ethnonyms mean ‘mountain people.’

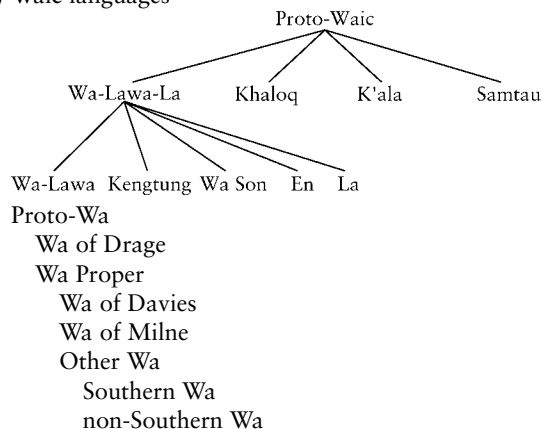
Members of the widespread Palaung-Wa branch of Mon-Khmer are found scattered throughout Myanmar, Thailand, the Yunnan province of China, and Laos. The total numbers of Palaung-Wa speakers is likely over one million. Several divergent groups are to be found within this branch, the exact internal relations between which remain to be worked out to the satisfaction of specialists. The major languages or subgroups within this branch are Danau, the various divergent Angkuic groups, Palaung proper, Riang, and the large Waic group with multiple subdivisions. Some reckon an Eastern and a Western group, the former including Danau, Palaung and Riang, the latter consisting of Waic, Angkuic and possibly Lametic as well.

Palaung speakers, who fall into several dialect/language groups, are heavily influenced by Shan and undergoing linguistic assimilation to this Tai language. At least three Palaung languages are reckoned, viz. Pale (Silver), Rumi, and Shwe (Gold). The Palaung languages are mainly spoken in Myanmar, but each have a small number of speakers in China, where they are all united under the official De'ang nationality; Pale Palaung speakers are also found in Thailand. The total number of Palaung speakers is difficult to estimate but may be in the range of 500 000–600 000, or it may be much smaller. A divergent Palaung group, the P'u-man, are found in Yunnan, China. As this Chinese ethnonym refers to other groups as well, it is not known how many speakers of P'u-man are really to be found. Riang, like Palaung, is heavily influenced by Shan and rapidly losing its similarly small population of speakers. Danau is also severely endangered and could have as few as 2000 speakers, if that many, at present.

Lametic, as mentioned above, shows considerable influence from Khmuic and may not really belong in this branch of Mon-Khmer. It consists of Con, with perhaps 1000 speakers, and Lamet, with around 10 000 total speakers or maybe more.

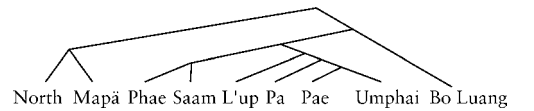
The large and diverse Waic languages of Palaung-Wa constitute a heterogeneous group of languages spoken in enclaves throughout Myanmar and Yunnan, and originally in adjacent parts of Thailand as well. The number of Waic languages and its internal divisions remain open questions, despite considerable work by Diffloth in particular. Most Waic languages are spoken by small populations which range from less than 100 to more than 100 000. Wa is often known as Va in China; other common ethnonyms referring to Wa-speaking groups include (Parkin, 1991: 111): Vu, Vo, Lave, Ravet, Krak, Kut Wa, Hsap Tai, and Gaung-pyat (head cutting). One possible subgrouping of this group is as follows:

(13) Waic languages

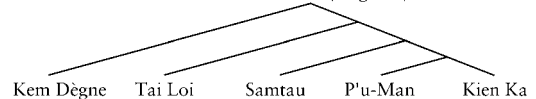


Kawa
non-Kawa
Tung Va Wa
Wa of Antisdel
'Bible Wa'
Praok

NB: Names refer to authors who described these languages.



Proto-Samtau (Augkuic)



NB: Samtauc/Angkuic P'u-Man is not the same as Palaungic P'u-Man.

As with other Mon-Khmer subgroups, there is a proliferation of names associated with languages of this group. Thus, to Angkuic may be found such names as Hu, Kiorr, Kon Keu, Man Met, Samtao, Tai Loi, and U. To Waic also belongs the official minority Bulang (Blang) language of China; this officially recognized ethnic designation in China also subsumes many other related languages of the Waic group. The Ethnologue reckons only four other Waic languages, Eastern Lawa, Western Lawa, Vo, and Parauk. As in many areas, the language/dialect distinction is ill defined and subject to the whims or biases of individual researchers.

The important languages of the Pearic branch of Mon-Khmer were spoken by around 8000–10 000 people in Cambodia before the ravages of the Vietnam War and the subsequent terror imposed by the Khmer Rouge regime. Only a handful of speakers of the half-dozen or so languages may remain. The languages of the Pearic branch include Chhong (Chong), known for its unusually developed system of register/voice quality contrasts characterizing its vowel system, Pear, Samre, Somray, Sa'och, and the poorly known Suoy (not to be confused with the Katuic-speaking group of the same name). Pearic peoples are dark skinned and have curly hair and as such were often discriminated against in Cambodia. Traditionally, the Pear proper were tribute payers of the Khmer in cardamom (Parkin, 1991: 68).

The large and diverse branch of Mon-Khmer known as Viet-Muong consists of an indeterminate number of languages spoken primarily in Vietnam and adjacent parts of Laos. First and foremost belonging to this branch is Vietnamese, far and away the Austro-Asiatic language with the most speakers, with perhaps

as many as 60–70 million. In fact, Vietnamese has more speakers than the other 150-odd Austro-Asiatic languages combined. Highly divergent within the family, with a developed tone system, lack of minor syllables and monosyllabic structure, lack of affixation processes, and heavy lexical influence from Chinese, the Austro-Asiatic affiliation of Vietnamese was not established until relatively recently (and is still disputed by some). Among the other languages of the branch, Muong stands out with its 400 000–500 000 speakers. Most other Viet-Muong languages have between several hundred and several thousand speakers. Most are poorly known or are indeed unattested linguistically, save perhaps an isolated word list. These are conventionally divided into a Chut subgroup, consisting of Arem, May, Pakatan, Ruc and Sach; a Cuoi subgroup, to which belong Hung, Pong, and Tum; a Muong group, consisting of Bo, Kha Tong Luang, Muong proper, Nguôn, and another Pong; a small Thavung-Phon Sung (Aheu) group; Vietnamese; and the poorly known and still unclassified Tho language. Other languages not recognized by the Ethnologue belonging to Viet-Muong include Coi, Dan Lai, K'katiam-Pong-Huok (Thai Pong), Ly Ha, Ma Lieng, Nguoi Rung, Nha Lang, Tay Cham, Tay Pum, and Tay Tum (Ktum) (Parkin, 1991). Many Viet-Muong languages are undergoing rapid assimilation to Vietnamese. While included here, it is possible that Viet-Muong is not actually a subgroup of Mon-Khmer, but, like Aslian and Nicobarese, a separate subgroup of (non-Munda) Austro-Asiatic.

In addition to the subgroups of Mon-Khmer languages adduced above, there are a number of as yet unclassified or isolated groups as well. Most of these are relatively recently described minority languages from China and Vietnam in particular.

The Mang or Mang U of Vietnam and China number perhaps 1000 speakers. Mang has similarities with Khmuic and Palaung-Wa, more with the latter, but may constitute its own subgroup within Mon-Khmer.

The Palyu, who occupy the Guangxi–Guizhou border region of China, have also been recently identified as a Mon-Khmer-speaking group, the exact affiliation of which remains to be demonstrated. They are locally known as Lai (not to be confused with the Tibeto-Burman Lai Chin of Bangladesh and Myanmar).

Other recently discovered and as yet unclassified Austro-Asiatic languages of China possibly belonging to Mon-Khmer include Bugan, Buxinhua, Kemiehua, and Kuanhua. What little is known of these languages suggests that they may well be important for comparative Mon-Khmer linguistics.

There is a specialist journal devoted to the linguistic analysis of Mon-Khmer languages, *Mon-Khmer*

Studies, where the interested reader may find a wide range of articles covering virtually every conceivable linguistic topic in the investigation of the languages of this family. The majority of articles are devoted to phonological analysis, as Mon-Khmer languages, as mentioned above, are particularly unusual in this domain. Among present-day specialists in Mon-Khmer linguistics, Gerard Diffloth deserves special mention.

Bibliography

- Anderson G D S & Zide N H (2002). 'Issues in Proto-Munda and Proto-Austroasiatic nominal derivation: the bimoraic constraint.' In Macken M A (ed.) *Papers from the 10th Annual Meeting of the Southeast Asian Linguistics Society*. Tempe: South East Asian Studies Program, Arizona State University. 55–74.
- Bauer C (1989). 'The verb in spoken Mon.' *Mon-Khmer Studies* 15, 87–110.
- Bauer C (1990). 'Reanalyzing reanalyses in Katuic and Bahnaric.' *Mon-Khmer Studies* 16–17, 143–154.
- Benjamin G (1976). 'Austroasiatic subgroupings and prehistory in the Malay peninsula.' In Jenner *et al.* (eds.). 37–128.
- Costello N A (1991). *Nôôq Paraaq Katu*. Manila: SIL Thailand.
- Diffloth G (1980). 'The Wa languages.' *Linguistics of the Tibeto-Burman Area* 5(2).
- Diffloth G (1984). *Dvaravati Old Mon and Nyah Kur*. Chulalongkorn, Thailand: Chulalongkorn University Press.
- Ferlus M (1979). 'Formation des registres et mutations consonantiques dans les langues mon-khmer.' *Mon-Khmer Studies* 8, 1–76.
- Gorgoniev Iu A (1974). *Glagol khmerskogo iazyka*. Moscow: Nauka.
- Gregerson K J (1976). 'Tongue-root and register in Mon-Khmer.' In Jenner *et al.* (eds.) 323–370.
- Grimes B (ed.) (2000). *Ethnologue: languages of the world* (14th edn.). Dallas: SIL International.
- Haudricourt A-G (1965). 'Mutation consonantique en mon-khmer.' *Bulletin de la Société Linguistique de Paris* 60, 160–172.
- Headley R K Jr (1977). 'A Pearic vocabulary.' *Mon-Khmer Studies* 6, 69–150.
- Huffman F E (1976). 'The register problem in fifteen Mon-Khmer languages.' In Jenner *et al.* (eds.). 575–590.
- Jacob J (1993). Smythe D A (ed.) *Cambodian linguistics, literature and history: collected articles*. London: School of Oriental & African Studies.
- Jenner P N & Saveros P (1982). *A lexicon of Khmer morphology*. Honolulu: University of Hawai'i Press.
- Jenner P N, Thompson L C & Starosta S (eds.) (1976). *Austroasiatic studies 1*. Honolulu: University of Hawai'i Press.
- Li J (1996). 'Bugan: a new Mon-Khmer language of Yunnan Province, China.' *Mon-Khmer Studies* 25, 135–160.
- Milne M L (1921). *Palaung grammar*. Oxford: Clarendon Press.

- Nagaraja K S (1985). *Khasi: a descriptive analysis*. Pune, India: Deccan College.
- Parkin R (1991). *A guide to Austroasiatic speakers and their languages*. Honolulu: University of Hawai'i Press.
- Paulsen D A (1992). 'Phonological reconstruction of Proto-Plang' *Mon-Khmer Studies* 18–19, 160–222.
- Peiros I (1996). *Katuic comparative dictionary*. Canberra: Pacific Linguistics.
- Schiller E (1990). 'On the definition and distribution of serial verb constructions.' In Joseph B & Zwicky A (eds.) *When verbs collide*. Columbus: Ohio State University Press. 34–64.
- Schmidt P W (1906). 'Die Mon-Khmer-Völker: ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens.' *Archiv für Anthropologie, Braunschweig n.s.* 5, 59–109.
- Shorto H L (ed.) (1963). *Linguistic comparison in South-east Asia and the Pacific*. London: School of Oriental & African Studies.
- Sidwell P & Jacq P (2004). *A handbook of comparative Bahnaric 1: West Bahnaric*. Canberra: Australian National University.
- Smith K D (1972). *A phonological reconstruction of Proto-North-Bahnaric*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Smith K D (1975). 'The velar-animal prefix relic in Vietnam languages.' *Linguistics of the Tibeto-Burman Area* 2(1), 1–18.
- Smith K D (1992). 'The-VC rhyme link between Bahnaric and Katuic.' *Mon-Khmer Studies* 18–19, 106–159.
- Sokolovskaia N K & Nguyen V T (1987). *Iazyk muong: materialy sovetsko-v'etnamskoi lingvisticheskoi èkspe-ditsii 1979 goda*. Moscow: Nauka.
- Svantesson J-O (1983). *Kammu phonology and morphology*. Lund: CWK Gleerup.
- Thomas D L (1971). *Chrau grammar*. Honolulu: University of Hawai'i Press.
- Thomas D & Headley R K (1970). 'More on Mon-Khmer subgroupings.' *Lingua* 25, 398–418.
- Zide N H & Anderson G D S (2001). 'The Proto-Munda verb: some connections with Mon-Khmer.' In Subbarao K V & Bhaskararao P (eds.) *Yearbook of South-Asian languages and linguistics 2001*. Delhi: Sage. 517–540.

Morphological Typology

C Iacobini, Università di Salerno, Salerno, Italy

© 2006 Elsevier Ltd. All rights reserved.

At the beginning of the 19th century, linguistic typology established a small set of types, i.e., isolating, agglutinating, fusional (see 'The Fusional Type,' 'The Agglutinating Type,' and 'The Isolating Type' below), to which any single language could be assigned. The main criteria for the assignment were related to word structure.

For a long time, word structure remained the dominant criterion to classify languages into types, so that morphological typology is sometimes defined as classical typology as well.

Friedrich Schlegel (1772–1829) and his brother August Wilhelm (1767–1845), together with Wilhelm von Humboldt (1767–1835) – who was also the first scholar who identified the polysynthetic type (see **Central Siberian Yupik as a Polysynthetic Language**) – gave the most important contribution to the assumption that it was possible to describe the whole grammatical structure of a language starting from the way in which relational concepts are morphologically encoded (following suggestions from contemporary research in botany and paleontology).

With Edward Sapir (1921) there was an important shift within morphological typology. Abandoning the holistic approach, he underlined the internal inconsistencies of the classical schema of classification, and

distinguished and made explicit the relevant parameters for classification. According to Sapir, it is possible (and usual) for the same language to show morphological structures belonging to more than one type.

Modern linguistic typology – which arose with the work of Joseph Greenberg (1966) – attempts to classify languages simultaneously on several dimensions, using implicational types of relations to establish limitations on the range of possible variation occurring within linguistic structures. Therefore, morphology is no longer seen as the most fundamental form of language classification, but one of the main levels in which languages can be described.

In contemporary typology, the classification by morphological types is mostly a convenient way to rapidly identify some morphological characteristic of languages, but it is marginal both in theoretical and descriptive studies.

If compared with those of the beginning of the 19th century, the aims of morphological typology have become more modest. However, the role played by morphology in the research of possible patterns of co-variation and correlation with other levels of linguistic analysis is still very relevant today.

Morphological Types

The classification of languages by morphological types is still today part of the standard terminology

of linguistics. However, it is also strongly criticized by the majority of typologists for three main reasons:

1. the classification criteria are rather vague and difficult to apply in a consistent way;
2. the morphological type is defined in terms of mutual favorability of properties rather than of implicational correlations, resulting in a low predictive power;
3. morphological typology has a holistic background.

Another reason for criticism comes from the emotive appeal of linguistic imperialism: modern linguistics has disavowed the ideological prejudice dating back to the beginning of the 19th century, according to which the fusional morphological type was considered superior to other types both on a functional and on an evolutionary scale (the presumed superiority stemmed from scholars' Western-centered standpoint: in fact, all the older and many of the contemporary Indo-European languages can be classified as fusional).

There are three fundamental conceptions of language type based on morphological criteria:

1. the classical (cf. Schlegel, 1808): each type is distinguished from the others in a clear-cut way and is characterized by the presence/absence of one single feature (e.g., languages with or without inflection);
2. the continuum (cf. Sapir, 1921): each type shares with the others a combination of various features, each of them having a continuum of values with two well-defined ends; the structures of a language can be placed on a specific point along the axis anchored by each feature;
3. the ideal (cf. Skalička, 1966): each type is an ideal model (which is never fully realized), consisting of a set of features which tend to co-occur.

According to the last conception, each morphological type may be described as a combination of functionally interconnected features, which, as a whole, form an ideal construct characterizing (the whole, or some aspects of) the morphology of languages (Sgall, 1995). Languages are rarely pure types; they usually mix elements of different types. Assigning a language to a specific type depends on the preponderance of features considered significant (the quantification of such features is a difficult problem to solve from a practical point of view).

Criticism of Morphological Types

The vagueness of the classification in morphological types is shown by the lack of consensus on the

number of both types and parameters identifying them. The three main types are: fusional, agglutinating, and isolating. There is no agreement either on whether the polysynthetic type should be considered autonomous or an extreme degree of the agglutinating one (*see Central Siberian Yupik as a Polysynthetic Language*), or on whether Semitic language features are sufficient to distinguish an introflexing type from the fusional one.

The classical morphological typology only referred to the formal encoding of single morphological features. Sapir (1921: Chap. 6) recognized the shortcomings and contradictions of the 19th century typologists, and, in modern linguistic terms, made explicit the formal components implicit in previous proposals. He adopted a multidimensional approach to morphological typology that could integrate different dimensions of classification, among which the semantic content of morphological expression as well. He distinguished three main criteria of classification:

1. morphological technique: isolating, agglutinating, fusional, symbolic (internal modification);
2. index of synthesis: analytic, synthetic, and polysynthetic structures;
3. how relational concepts are expressed (making a distinction on whether they are expressed through lexical bases or relational elements), and the degree of grammaticalization of relational concepts.

With Sapir, there is a move from a taxonomic criterion according to which each language has to be ascribed to one type, to a classifying criterion based on possible types of morphological structures, in the definition of which semantics plays an important role as well. Moreover, it is explicitly recognized that a language can be classified into more than one type, and that types shade one into the other.

Greenberg (1954) elaborated Sapir's proposal with the aim of applying it to quantitative evaluations on the morphology of different languages.

The three parameters that are nowadays mostly used for the typological classification of languages are the ratio of morphs to word forms, the number of morphemes to morphs, and the degree of word-internal modification of morphs. The first parameter distinguishes analytic from synthetic languages; the other two distinguish, within the synthetic group, agglutinating from fusional languages.

Lists of Clustering Features

The best-known attempt to establish a list of features that co-occur in morphological types is the one made by Skalička (1966), which also includes aspects of

word phonology and word order. In the following sections, the features of the three main types (agglutinating, fusional, and isolating) will be given (*see also Finnish as an Agglutinating Language; Italian as a Fusional Language; Chinese as an Isolating Language*), with the reminder that whereas the final ends of some dimensions can be reached, others are unlikely or impossible to reach. The analysis and discussion of polysynthetic and introflexing types is carried out in sections devoted to them (*see Central Siberian Yupik as a Polysynthetic Language; Arabic as an Introflexing Language*). Suffice it to say here that whereas the polysynthetic type presents a variety of forms and is spread in a lot of languages, the introflexing type is not widely spread and, even in those languages where it is found, it is restricted to a part of morphology.

The features that tend to cluster in languages displaying one of the three main morphological types can be listed as shown in the following sections.

The Fusional Type

1. Words are formed by a root and (one or more) inflectional affixes, which are employed as a primary means to indicate the grammatical function of the words in the language. Agreement is widely employed.
2. High degree of modification of internal morph boundaries, with a consequently difficult linear segmentation.
3. Tendency to cumulate morphological meanings in a single affix (with consequent asymmetry between the semantic and formal organization of grammatical markers).
4. Word-class distinction is maximal. Inflection is rich, as regards both the number of inflectional classes and the extension of paradigms.
5. Stem suppletion; many cases of both homonymy and synonymy among affixes; clear distinction between inflectional and derivational affixes.
6. A slight correlation with syntax can be seen in the relatively free word order (but there are also fusional languages with a fairly fixed word order).

The Agglutinating Type

1. Words are formed by a root and a clearly detachable sequence of affixes, each of them expressing a separate item of meaning. Affixes are widely employed to indicate the relationships between words. Therefore, there are few or no independent relational elements (e.g., pronouns, pre-/postposition, articles, etc.), and a wide use of nominal cases.

2. Very high matching between morphs and morphemes. Morphs are loosely joined together; consequently it is very easy to determine the boundaries between them.
3. Each affix carries only one meaning; no cases of homonymy or synonymy among affixes; the semantic structure is directly reflected in the morphological articulation of the word; no principled limits to the number of affixes in a word.
4. Word-class distinction is minimal: the same affixes tend to occur with roots belonging to different parts of speech (e.g., personal endings to nouns, case endings to verbs); almost the same morphology for adjectives and verbs. No inflectional classes, no gender distinction.
5. Derivational affixes are widely employed in word formation. The distinction between inflectional and derivational affixes is slight. Many affixes reveal their lexical origin to some extent. The latter feature, together with the tendency of affixes to form autonomous syllables and to be relatively unconstrained in number, results in words that are quite long.
6. Relatively fixed word order. Agreement is almost completely absent.

The Isolating Type

1. Words are monomorphemic, invariable, and formed by a single root. Ideally, bound forms are completely missing. Position is the main way of expressing the relationship between independent words.
2. Relational meanings are not overtly expressed, or the same units that normally encode lexical concepts are used for that purpose as separate helper words; the meaning and function of a word considerably depend upon the syntagmatic context.
3. There is little to no morphological complexity. Morphs are clearly identifiable both phonologically and semantically: morph boundaries are sharply defined, phonological form is invariant, there are no instances of overlapping exponence. Derivation is nonexistent, partly replaced by compounding.
4. The distinction in parts of speech is not clear; there is no overt expression of grammatical categorization.
5. Tendency to monosyllabism with no phonetic distinctions between the elements expressing lexical meaning and the ones expressing relational meaning.
6. Rigid word order.

Comparison

The fusional type is differentiated from the isolating type by the use of bound morphs and the clear-cut distinction between word classes; it is differentiated from the agglutinating type by the kind of juncture between morphs, and the nonbiunivocal correspondence between morphs and morphemes. In the synthetic vs. analytic distinction, the fusional and even more the agglutinating type tend toward the synthetic end.

Types characterization can be schematically summed up in Table 1.

Contemporary Morphological Typology

In spite of criticism, the classification by morphological types is still convenient and widely used in order to rapidly identify a number of features that tend to co-occur in the morphology of a language. Some authors also use it to assess the extent to which a language moves away from such ideal constructs both in a synchronic and in a diachronic perspective (Dressler, 1985 argues that languages tend to move toward a typological goal according to a linguistic economy criterion).

On the contrary, the classification by morphological types is now negligible in contemporary typological research, since scholars do not expect the morphological type to correlate in a significant way with other typological parameters. As they refuse the possibility of classifying the whole of a language into a given type, typologists pay increasing attention to the practice of partial typology, focusing on specific areas of linguistic structure (Bynon, 2004). Partial typology analyzes clusters of properties with a view to ascertaining significant connections and hierarchical organization. The main tool for analysis is establishing implication universals according to Greenberg's (1966) suggestions. However, disregarding morphological types does not result in the death

of morphological typology: almost half of Greenberg's 45 universals concern morphology. These universals mainly focus on two aspects previously neglected by morphological typology:

1. the relative order in which (derivational and inflectional) concepts are expressed morphologically within word forms;
2. the hierarchy of concepts which a language expresses morphologically.

Examples of the first are universals 28 and 39. The former states that if derivational and inflectional affixes are on the same side of the root, then the derivation is always closer to the root. The latter says that in a noun the expression of number is nearly always closer to the base than the expression of the case. Examples of the second are universal 36 'If a language has the category of gender, it always has the category of number,' and 37 'A language never has more gender categories in nonsingular numbers than in the singular.'

While it is true that only with Greenberg (1966) does syntax start to play a major role in typology, it is to be said that even the famous syntactic universals concerning word order (e.g., universals 3 and 4, which relate VSO order with preposition, and SOV order with postposition) show a link with morphology, for example, via universal 27 (which relates suffixing and postpositional languages on the one hand and prefixing and prepositional languages on the other). Another example is universal 41, which states the implication between the SOV word order and the presence of a case system. Because of such a link, morphology has acquired importance for scholars interested in word order typology. There are many studies that associate the use of suffixes or prefixes with OV and VO order, respectively (see Cutler *et al.*, 1985; Hawkins and Cutler, 1988; Hawkins and Gilligan, 1988, where affix preference is related with head position and also explained by

Table 1 Features characterizing fusional, agglutinating, and isolating types

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Fusional | Yes | No | No | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | No | Yes |
| Agglutinating | Yes | Yes | Yes | No | No | Yes | No | Yes | Yes | No | No | Yes | No | No |
| Isolating | No | Yes | Yes | No | Yes | Yes | Yes | No | No | No | No | No | Yes | No |

(1) affixes; (2) word function determined by position; (3) possibility of segmentation (morph-morpheme correspondence); (4) clear-cut distinction in part-of-speech; (5) tendency to monosyllabism; (6) fixed syntactic order; (7) syntagmatic structures employing particles; (8) noun marked for number; (9) noun marked for case; (10) noun marked for gender; (11) adjectives expressing agreement; (12) synthetic expression of comparison on adjectives; (13) verb: use of analytic structures (vs. affixes); (14) verb: presence of inflectional classes.

psycholinguistic factors; see Bybee *et al.*, 1990 for a diachronic approach; Dryer, 1992, who attempts to integrate different approaches).

At the interface between morphology and syntax there are works such as Nichols (1986), which distinguishes head- and dependent-marking, and others, such as those dealing with the morphologically based typology of causative constructions.

Greenberg (1966) barely takes into account phonology. However, the Universals Archive (cf. Plank and Filimonova) collects some universals concerning the interaction between morphology and phonology. Examples are universal 219 stating that affixes have a more limited inventory of phonemes than roots and universal 713, which correlates agglutinative morphology with vowel harmony, and fusional morphology with stress accent.

As far as morphology itself is concerned, the main effect of Greenberg's typology has been to stimulate the investigation of implicational relations between morphological categories, to promote the study of the markedness (overt expression) of morphological values, as well as the investigation of the reasons for the greater or lesser closeness of relational morphemes to the lexical base, and of the role played by morphological heads.

One of the main problems that stands before morphological typology is to identify a phenomenon as the same thing in different languages, since the values of one grammatical category may be expressed morphologically in some languages and in a different way in others. Cross-linguistic identification on purely formal criteria is not far-reaching (an example is Greenberg' universal 26: 'If a language has discontinuous affixes, it always has either prefixing or suffixing or both'); on the other hand, purely functional statements are not sufficient, because, by definition, they need a formal counterpart to be morphological. This type of difficulty explains why studies on inflectional morphology are overwhelmingly prevailing over derivational ones: categories such as case, gender, and number are cross-linguistically better identified and defined (see Blake, 2001; Corbett, 1991, 2000) than derivational ones; furthermore, the internal articulation of inflectional categories, featuring a number of values within one category (e.g., singular, plural for number), favors cross-linguistic comparison within the inflectional domain. The study of the variation in the expression of derivational categories is certainly more difficult and much remains to be done; see Bauer (2002) for a recent attempt.

Bibliography

- Bauer L (2002). 'What you can do with derivational morphology.' In Bendjaballah S *et al.* (eds.) *Morphology 2000. Selected papers from the 9th morphology meeting*. Amsterdam: Benjamins.
- Blake B J (2001). *Case* (2nd edn.). Cambridge: Cambridge University Press.
- Bybee J L, Pagliuca W & Perkins R D (1990). 'On the asymmetries in the affixation of grammatical material.' In Croft W, Denning K & Kemmer S (eds.) *Studies in typology and diachrony*. Amsterdam: Benjamins.
- Bynon T (2004). 'Approaches to morphological typology.' In Booij G, Lehmann C & Mugdan J (eds.) *Morphologie – morphology*, vol. 2. Berlin: De Gruyter.
- Corbett G (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett G (2000). *Number*. Cambridge: Cambridge University Press.
- Cutler A, Hawkins J A & Gilligan G (1985). 'The suffixing preference: a processing explanation.' *Linguistics* 23, 723–758.
- Dressler W U (1985). 'Typological aspects of natural morphology.' *Acta Linguistica Academiae Scientiarum Hungaricae* 35, 51–70.
- Dryer M S (1992). 'The Greenbergian word order correlations.' *Language* 68, 81–138.
- Greenberg J H (1954). 'A quantitative approach to the morphological typology of language.' In Spencer R F (ed.) *Method and perspective in anthropology*. Minneapolis: University of Minnesota Press.
- Greenberg J H (1966/1963). 'Some universals of grammar with particular reference to the order of meaningful elements.' In Greenberg J H (ed.) *Universals of language* (2nd edn.). Cambridge: MIT Press.
- Hawkins J A & Cutler A (1988). 'Psycholinguistic factors in morphological asymmetry.' In Hawkins J A (ed.) *Explaining Language Universals*. Oxford: Blackwell.
- Hawkins J A & Gilligan G (1988). 'Prefixing and suffixing universals in relation to basic word order.' *Lingua* 74, 219–259.
- Nichols J (1986). 'Head-marking, and dependent-marking grammar.' *Language* 62, 56–119.
- Plank F & Filimonova E (eds.) *The universals archive*, website <http://ling.uni-konstanz.de/591/Universals/introduction.html>
- Sapir E (1921). *Language: an introduction to the study of speech*. New York: Harcourt, Brace & World.
- Schlegel F (1808). *Ueber die Sprache und Weisheit der Indier: Ein Beitrag zur Begründung der Altertumskunde*. Heidelberg: Mohr and Zimmer.
- Sgall P (1995). 'Prague School typology.' In Shibatani M & Bynon T (eds.) *Approaches to language typology*. Oxford: Clarendon Press.
- Skalička V (1966). 'Ein "typologisches Konstrukt."' In Skalička V (ed.) (1979). *Typologische Studien*. Braunschweig: Vieweg.

Morrobalama

S Ogilvie, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Morrobalama (Umbuygamu) is an endangered Australian Aboriginal language of the Cape York Peninsula in northern Australia. Originally spoken in Princess Charlotte Bay on the eastern coast of Cape York, its speakers were forcibly displaced from the region in the early 1960s. A handful of remaining speakers now live in the Cape York towns of Umagico and Coen. Morrobalama is atypical of most Australian languages in its phonemic inventory but typical in all other respects: it is nonprefixing with nominal case inflections, verbal inflections, and pronominal cross-referencing. It has free word order with a complex verbal structure that includes a fixed order of verbal stems, tense/aspect markers, and pronominal clitics.

With the other languages of the Princess Charlotte Bay region – Lama-Lama, Umbindhamu, Rimanggudinhma (Mbariman-Gudhinma), and the Flinders Island language – Morrobalama forms a subgroup of the northern Paman languages known by some linguists as the ‘Lamalamic’ languages or ‘Bay Paman.’ While speakers of the language refer to it as ‘Morrobalama,’ it is also called ‘Umbuygamu’ in the neighboring Umpila language, and this name had been used by linguists for many years. ‘Morrobalama’ is now the name preferred by both speakers and linguists (see Dixon, 2002: xxxi).

Australian languages are known to be relatively homogeneous in their phonemic inventories. Typically they display four to six paired stops and nasals (labial, apico-alveolar, apico-postalveolar, lamino-dental, lamino-palatal, and dorso-velar), two rhotics (trill and retroflex), no fricatives, no voicing contrasts, and a three-vowel system. In contrast, Morrobalama has a large phonemic inventory that includes atypical sounds such as fricatives, prestopped nasals, voicing contrasts, and a system of five vowels that contrast in length. More thorough analysis is needed of the sound system, but it may have seven places of articulation, including two laminal series (dental and palatal), three rhotics (two of which contrast in voice), two glides, and a glottal stop which is only found in a few other Australian languages (in western Australia and in Arnhem Land). While a typical Australian inventory consists of seventeen phonemes, the Morrobalama inventory has expanded to include thirty-six phonemes.

Morrobalama and other languages of Cape York have undergone phonological changes that may at

first glance make them appear unrelated to other Australian languages. One such well-described change is the loss of the initial consonant (in the case of Morrobalama) or the initial syllable (in the case of other Princess Charlotte Bay languages). Thus, for example, **nyura* ‘you (all)’ has become *orha* and **nyulu* ‘s/he’ has become *ola*. Initial dropping also occurs in other languages in separate parts of Australia, such as Nhanda in western Australia and Arrernte in central Australia.

Morphologically, Morrobalama is typical of most Australian languages in that it displays a split-ergative system: nouns operate in an absolutive/ergative paradigm, while pronouns are nominative/accusative. Thus in Morrobalama, absolutive is marked by zero on both the subject of an intransitive verb and the object of a transitive verb, and the ergative is marked by *-a* suffixed to the subject of a transitive verb. Pronouns have three numbers – singular, dual, and plural – and distinguish inclusive and exclusive in first-person dual and plural. They can occur both independently or bound, and when bound, occur as the final constituent of the verbal structure with the initial vowel dropped. So the independent pronoun *ola* ‘s/he’ becomes bound as *-la*:

| | |
|-----------------------|---------------------------------------|
| <i>ola</i> | <i>nya – la – nan</i> |
| 3RD.SING.ACC | hit-3RD.SING.ACC. she-2ND. SING. OBJ. |
| ‘she (emph.) hit you’ | |

This example also shows a typical Australian word-order feature: order is generally free, but there is a tendency for it to be based on pragmatic rather than grammatical principles, in which the emphasized phrase occurs first. In Morrobalama, all verbs must include at least a verbal stem plus a pronominal subject clitic; order of the clitics is free. Other suffixation on the verb marks tense, aspect, and mood (TAM). There is fixed order within the verbal structure (stem-TAM-PRO).

As is typical in Australian languages, Morrobalama derives new words via suffixes. For example, the comitative case *-pinh* acts as a lexical formative by being suffixed to a word to form a compound *x-pinh*, meaning ‘with or having x’:

| | | |
|------------------|---------------------|------------------------|
| <i>ithi-pinh</i> | <i>lirrin-pinha</i> | <i>marr-pinh</i> |
| bone-COM | smoke-COM | face-COM |
| (‘with bone’) | (‘with smoke’) | (‘with face’) |
| ‘bony bream’ | ‘steam boat’ | ‘red-faced flying fox’ |

Most compounds, though, are of the ‘general-specific’ type, with stress falling on the specific element as though the general component acted as a clitic:

| | |
|--------------------|--------------------|
| lam-eethal | lam-agaparr |
| hand-bone | hand-belly |
| 'back of the hand' | 'palm of the hand' |

Morrobalama has not been passed on to the next generation, which speaks Cape York Creole (Torres Strait Creole) instead and, given the age of the few current speakers, it will probably disappear in the next decade unless intense language revitalization takes place.

Bibliography

- Dixon R M W (1980). *The languages of Australia*. Cambridge: Cambridge University Press.
 Dixon R M W (2002). *Australian languages: their nature and development*. Cambridge: Cambridge University Press.
 Dixon R M W & Blake B J (eds.) (1979, 1981, 1983, 1991, 2000). *The handbook of Australian languages* (5 vols).

vols 1 to 3, Canberra: Australian National University Press and Amsterdam: John Benjamins; vols 4 and 5, Oxford: Oxford University Press.

- Ogilvie S (1994). *A wordlist of the Morrobalama (Umbuygamu) language of Cape York, Australia*. Canberra: Australian National University Press.
 Ogilvie S (1994). *The Morrobalama (Umbuygamu) language of Cape York Peninsula, Australia*. M.A thesis, Australian National University.
 Rigsby B (1992). 'The languages of the Princess Charlotte Bay region.' In Dutton T, Ross M & Tryon D (eds.) *The language Game: papers in memory of Donald C. Laycock*. Canberra: Research School of Pacific Studies, Australian National University. 353–360.
 Sommer B (1976). 'Umbuygamu: the classification of a Cape York Peninsula language.' *Papers in Australian Linguistics* no. 10 *Pacific Linguistics A* 47, 13–30.
 Sutton P (ed.) (1976). *Languages of Cape York*. Canberra: Australian Institute of Aboriginal Studies.

Munda Languages

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

'Munda' is a group of languages belonging to the Austroasiatic language family spoken in eastern and central India, primarily in the states of Orissa, Jharkand, Bihar, and Madhya Pradesh and in adjacent areas of West Bengal, Maharashtra, and Andhra Pradesh. Some Munda speakers are found in expatriate or diaspora communities throughout India, Nepal, and western Bangladesh as well. In earlier literature, Munda is often referred to as Kol or Kolarian.

The Munda language family recognizes a major split between a North Munda and a South Munda subgroup. Within the North Munda subgroup, there is a binary opposition between Korku and a large group of Kherwarian languages, which is perhaps more properly described as a dialect continuum. Kherwarian includes both the largest of the Munda languages, Santali, with nearly 7 million speakers, as well as some of the smallest, such as Birhor, with under 1000. Other noteworthy North Munda languages include Mundari and Ho, each with approximately 1 million speakers, and smaller languages such as Agariya, Asuri, Bhumij (Mundari), Karmali Santali, Koraku, Korwa, Mahali, and Turi. Publications may be found in the larger of the Kherwarian languages (Mundari, Ho, Santali), including a range of Santali publications in a native orthography (the Ol'

Cemet script). Short wave radio broadcasts are also available in Santali. The newly founded 'tribal' state of Jharkhand has a Munda-speaking majority and is lobbying to have some form of Kherwarian declared another state language of India.

The South Munda subgroup is older and more internally diversified than North Munda. At least the following languages belonging to this subgroup: Sora (Savara), Juray, Gorum (Parengi, Parenga), Gutob (Gadaba, Bodo), Remo (Bonda, Bondo), Gta? (Gata?, Didey), Kharia, and Juang. In terms of further subgrouping, it is clear that Sora and Gorum form a branch of their own, as do the closely related Gutob and Remo. Gta?, which is properly speaking two separate languages: the poorly known Hill Geta? and Plains/Riverside Geta?, has been traditionally linked at a slightly higher taxonomic level with Gutob-Remo (so-called Gutob-Remo-Gta?), and Kharia and Juang have been linked together in a branch as well. These latter two classifications are tenuous and remain to be adequately demonstrated (Anderson, 2001). South Munda languages range in total number of speakers from 300 000 or more Sora speakers to between 150 000 and 200 000 Kharia speakers to Gutob with approximately 30 000–50 000 total speakers and Juang with around 15 000 speakers. The remaining South Munda languages have around 2000–4000 speakers each.

Typologically speaking, all Munda languages are moderately to extensively agglutinating, show SOV basic clause structure, and possess preglottalized or

unreleased ‘checked’ consonants. This latter feature is quintessentially Munda, and readily distinguishes Munda languages from the surrounding languages of the subcontinent.

The shift to SOV word order from SVO to VSO may be attributed to influence from Indo–Aryan or Dravidian languages (Anderson, 2003), which varies from moderate to strong in individual Munda languages. Generally speaking, the southernmost South Munda languages show the greatest degree of structural influence from local tribal Dravidian and lexical influence from the local tribal Indo–Aryan (e.g., Desia/Kotia Oriya). In addition, Kharia has been heavily influenced by Kherwarian North Munda languages such as Mundari.

Kherwarian North Munda languages are characterized by a high degree of morphological complexity, standing out even within the morphologically rich languages of South Asia. Among the most noteworthy characteristic features of Kherwarian verb structure is the extensive system of referent indexing that includes, subjects, objects (direct and indirect), benefactives, and possessors as well. An example of this type of marking may be seen in the following Santali form:

- (1) Santali (North Munda)
sukri-ko gɔːc-ke-d-e-tiñ-a
 pig-PL die-ASP-TR-3-1POSS-FIN
 ‘they killed my pig’
 (Bodding, 1929: 209)

South Munda languages show a lesser degree of morphemic complexity than their sister languages to the north. Making this assertion is not to say that large complexes are not found among the South Munda languages, however. All South Munda languages reflect some degree of noun incorporation, either as an active morphosyntactic process (Sora, Gtaʔ) or preserved only in lexicalized formations (Kharia). In fact, Sora is among the very small number of the world’s languages that allows for both instances of multiple noun incorporation (2), as well as the typologically unusual agent incorporation with transitive verbs (3).

- (2) Sora (South Munda)
jo-me-bob-dem-te-n-ai
 smear-oil-head-RFLXV-NPAST-INTR-CLOC/1
 ‘I will anoint my head with oil’
 (Ramamurti, 1931: 143)
- (3) Sora (South Munda)
- | | |
|------------------------|------------------------|
| <i>ñam-kit-t-am</i> | <i>sa-bud-t-am</i> |
| seize-tiger-NPAST-2 | mangle-bear-NPAST-2 |
| ‘tiger will seize you’ | ‘bear will mangle you’ |
| (Ramamurti, 1931: 40) | (Ramamurti, 1931: 142) |

South Munda languages also show agreement in the verb with a possessor of a logical argument, but unlike Santali, which utilizes a special set of possessive inflectional suffixes, South Munda languages like Gorum show the more typologically ‘normal’ pattern of ‘Possessor Raising’ (4), raising the possessor to term argument and encoding it in the verb in a manner identical to object marking.

- (4) Gorum (South Munda)
putiputi-nom ir-om luʔr-om
 heart-2 beat-2 AUX-2
 ‘your heart is beating’
 (Aze, 1973)

Among the more salient phonological features of the Munda languages from a South Asia areal perspective include, in addition to the characteristic ‘checked’ consonants mentioned above (5), are such areally atypical features of individual Munda languages as low tone in Korku, creaky voice in Gorum, prenasalized stops in Santali, and, unusual syllable structure in Gtaʔ following a series of regular sound changes (seen in such words as Gtaʔ itself).

Munda languages also make extensive use of auxiliary verb constructions. Both a wide range of formal and functional subtypes of auxiliary verb constructions are found across the Munda language family. In general, these are in keeping with both typological and areal norms. Thus, one finds the use of the verb ‘eat’ in a passive construction found in a number of South Asian languages in such languages as Juang (6). Further, doubled or serialized agreement is found in auxiliary verb constructions in Gorum (7), and is characteristic of local Dravidian languages as well (Parji, Gondi, etc.).

- (6) Juang
- | | |
|--------------------------|--------------------------|
| <i>aiñ ma’d-jim-seke</i> | <i>aiñ ma’d-jim-sero</i> |
| I beat-eat- | I beat-eat-PERF.PAST.T |
| PERF.PRES.T | |
| ‘I am beaten’ | ‘I was beaten’ |
| (Pinnow, 1960) | |
- (7) Gorum
miñ ne-gaʔ-ru ne-laʔ-ru
 I 1-eat-PAST 1-hit-PAST
 ‘I ate vigorously’
 (Aze, 1973)

Among present-day researchers of Munda languages, Norman Zide deserves special mention.

Bibliography

Anderson G D S (2001). ‘A new classification of Munda: evidence from comparative verb morphology.’ *Indian Linguistics* 62, 27–42.

- Anderson G D S (2003). 'Dravidian influence on Munda.' *International Journal of Dravidian Linguistics* 32(1), 27–48.
- Anderson G D S & Boyle J P (2002). 'Switch-reference in South Munda.' In Macken M A (ed.) *Papers from the 10th Annual Meeting of the Southeast Asian Linguistics Society*. Tempe, AZ: Arizona State University, South East Asian Studies Program, Monograph Series Press. 39–54.
- Anderson G D S & Zide N H (2001). 'Recent advances in the reconstruction of the Proto-Munda Verb.' In Brinton L (ed.) *Historical linguistics 1999*. Amsterdam: Benjamins. 13–30.
- Aze F R (1973). 'Clause patterns in Parengi-Gorum.' In Trail R L (ed.) *Patterns in clause, sentence, discourse in selected languages of India and Nepal 1*. Norman, OK: SIL. 235–312.
- Bhat D N S (1997). 'Noun-verb distinction in Munda languages.' In Abbi A (ed.) *Languages of tribal and indigenous peoples of India: the ethnic space*. Delhi: Motilal Banarsidass. 227–251.
- Bhattacharya S (1966). 'Some Munda etymologies.' In Zide N H (ed.) *Studies in comparative Austroasiatic linguistics*. The Hague: Mouton. 28–40.
- Bhattacharya S (1969). 'Some more Munda etymologies.' In Heesterman J C, Schokker G H & Subramoniam V I (eds.) *Pratidaanam, Indian, Iranian and Indo-European studies presented to F. B. J. Kuiper*. The Hague. 362–370.
- Bhattacharya S (1972). 'Dravidian and Munda: a good field for areal and typological studies.' In *Third seminar on Dravidian linguistics*. Annamalai University. 241–265.
- Bhattacharya S (1975a). *Studies in comparative Munda linguistics*. Simla: Indian Institute of Advanced Study.
- Bhattacharya S (1975b). 'Linguistic convergence in the Dravido-Munda culture area.' *International Journal of Dravidian Linguistics* 4, 199–214.
- Bodding P O (1929). *Materials for a Santali grammar, mostly morphological*. Dumka: Santal Mission.
- Donegan P (1993). 'Rhythm and vocalic drift in Munda and Mon-Khmer.' *Linguistics of the Tibeto-Burman Area* 16(1), 1–43.
- Hook P (1991). 'The compound verb in Munda: an areal and typological overview' In Abbi A (ed.) *India as a Linguistic Area Revisited. Special issue of Language Sciences* 13(2), 181–195.
- Konow S (1904). 'Mundas and Australians.' *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 58, 147–157.
- Osada T (1992). *A reference grammar of Mundari*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Osada T (1996). 'Notes on the Proto-Kherwarian vowel system.' *Indo-Iranian Journal* 39, 245–258.
- Pinnow H-J (1960). *Beiträge zur Kenntnis der Juang-Sprache*. Unpublished manuscript.
- Pinnow H-J (1963). 'The position of the Munda languages within the Austroasiatic language family.' In Shorto H L (ed.) *Linguistic comparison in Southeast Asia and the Pacific*. London: SOAS. 140–152.
- Pinnow H-J (1966). 'A comparative study of the verb in the Munda languages.' In Zide N H (ed.) *Indo-Iranian Monographs 5: Studies in comparative Austroasiatic linguistics*. The Hague: Mouton. 96–193.
- Ramamurti G V (1931). *A manual of the So:ra: (or Savara) language*. Madras: Government Press.
- Stampe D L & Zide N H (1968). 'The place of Kharia-Juang in the Munda family.' In *Indian Linguistics Emeneau Sastipurti Volume*. Poona: Linguistic Survey of India. 370–377.
- Zide N H (1976). '"3" and "4" in South Munda.' *Linguistics* 174, 89–98.
- Zide N H (1978). *Studies in the Munda numerals*. Mysore: Central Institute of Indian Languages.
- Zide N H (1985). 'Notes mostly historical on some participant roles in some Munda languages.' In Zide A R K, Magier D & Schiller E (eds.) *Proceedings of the Conference on Participant Roles: South-Asia and Adjacent Areas*. 92–103.
- Zide N H (1999). 'Three Munda scripts.' *Linguistics of the Tibeto-Burman Area* 22, 199–232.
- Zide N H & Zide A (1976). 'Proto-Munda cultural vocabulary: evidence for early agriculture.' In Jenner P N, Thompson L C & Starosta S (eds.) *Oceanic Linguistics, Special Publication 13: Austroasiatic studies 2*. Honolulu: University of Hawaii. 1295–1334.

Muskogean Languages

G A Broadwell, State University of New York, Albany, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Muskogean languages were originally spoken across much of the southeastern United States, from Georgia to Louisiana. As a result of population movements, both voluntary and forced, many speakers of

Muskogean languages are now located outside the original Muskogean-speaking area. More than half of the speakers of Muskogean languages now live in either Oklahoma or southern Florida.

There are six different Muskogean languages that are still spoken.

1. Choctaw is the largest Muskogean language, with perhaps 7000 to 11 000 speakers divided between the Mississippi Choctaw reservation in eastern Mississippi and the Oklahoma Choctaw nation of southeastern Oklahoma.

2. Chickasaw is spoken by a few hundred people in the Chickasaw nation of southern Oklahoma.
3. Alabama is spoken by a few hundred people on the Alabama-Coushatta reservation in eastern Texas.
4. Koasati (also called Coushatta) is spoken by a few hundred people in the area around Elton, Louisiana, and by a smaller number (probably fewer than 50) on the Alabama-Coushatta reservation.
5. Mikasuki (also spelled Miccosukee) is spoken by approximately 1000 people in the Seminole and Miccosukee tribes of Florida, located in the Everglades region of Florida.
6. Creek (also called Muskogee or Muscogee) is spoken by approximately 4000 people. The great majority of these speakers live in the Muscogee (Creek) Nation of Oklahoma and in the Seminole nation of Oklahoma. There are perhaps 200 speakers of Creek in the Seminole tribe of Florida.

'Seminole' is a term that is potentially confusing. It refers primarily to a political grouping of Creek and Mikasuki speakers who moved from their former territory in Georgia and Alabama to new locations in south Florida beginning in the mid-18th century. During the Indian removal period of the 1830s, many Seminoles were forcibly resettled in Oklahoma, so that there is now both a Seminole Nation of Oklahoma and a Seminole tribe of Florida.

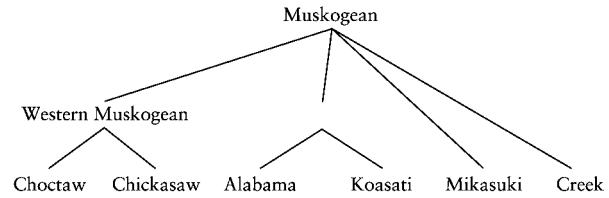
Oklahoma Seminoles who retain their native language speak a dialect of Creek called Oklahoma Seminole Creek. Florida Seminoles speak two different languages – the majority speak Mikasuki and a minority speak a dialect known as Florida Seminole Creek.

In addition to these six languages, three other extinct Muskogean languages are attested:

7. Apalachee was spoken by inhabitants of north-west Florida. The language is attested in a 17th century letter to Charles II of Spain, but has long been extinct.
8. Hitchiti was a language that seems to have been quite similar to Mikasuki. It was spoken in Florida and Oklahoma until the early 20th century.
9. Mobilian Jargon (Mobilian) was a Muskogean-based trade language spoken in the lower Mississippi Valley. Some fragments of the language were spoken by a few people in Louisiana until the 1970s.

Classification

The currently spoken Muskogean languages fall into the following groups:



Kimball and Haas have argued that the extinct language Apalachee was most closely connected to Alabama and Koasati.

Higher level classification of the Muskogean languages is difficult and a subject of controversy. Based on several identifiable sound changes, Haas proposed that Alabama, Koasati, Mikasuki, and Creek formed a group called Eastern Muskogean. However, another subsequently discovered sound change (Proto-Muskogean $*k^w \rightarrow b$) supports a rather different grouping, in which Western Muskogean, Alabama, Koasati, and Mikasuki form a group called Southern Muskogean (affected by the $*k^w \rightarrow b$ rule). This group is distinct from Creek, which was not affected by this rule. Munro and Broadwell have also presented morphological evidence in favor of a Southern Muskogean group. There is as yet no consensus on this issue.

Phonological Characteristics

Muskogean languages have relatively simple phonological inventories. Choctaw and Chickasaw, for instance, have an inventory of 16 consonant phonemes (Table 1).

In addition, all Muskogean languages have three phonemic vowels /a, i, o/ that appear short, long, and nasalized.

Muskogean languages have pitch-accent. In verbs, the position of pitch accent is generally dependent on the 'grade' in which the verb appears. Grades are a series of related verb forms that differ in tense and/or aspect from each other, and that are formed by infixing a consonant or consonants, lengthening a vowel, or

Table 1 Choctaw and Chickasaw consonant phonemes

| Bilabial/ Labiodental | Alveolar | Postalveolar/ Palatal | Velar | Glottal |
|--------------------------|----------|--------------------------|-------|---------|
| p | t | tʃ | k | ʔ |
| b | | | | |
| f | s | ʃ | | h |
| m | n | | | |
| | l | | | |
| | ɬ | | | |
| w | | j | | |

/ʔ/ in most Choctaw dialects is restricted to word-final position, while it has a wider distribution in Chickasaw.

changing the form and position of the pitch accent. Consider the following examples from Creek:

- (1) Creek
 wanáj-as [wanáyas] 'tie it'
 underived grade
 wana<:>j-is [wana:jís] 's/he is tying it'
 lengthened grade
 waná<h>j-is [wanáhjis] 's/he tied it (today, last night)'
 aspirated grade
 wanâ<:>j-is [wanâ:jís] 's/he has tied it.'
 falling-tone grade
 wanã<:~>j-is [wanã:jís] 's/he keeps tying it.'
 nasalizing grade

In these examples, the lengthened grade is associated with eventive aspect, the aspirated grade with the perfective aspect, the falling tone grade with the stative perfective aspect, and the nasalizing grade with the expressive grade (a grade expressing a large degree of something, sometimes equivalent to a continuative).

The infixed material or vowel length change in a grade appears in the syllable that is historically the penult of the verb stem. This continues to be the location of infixation and lengthening in Western Muskogean, Alabama, and Koasati. Due to loss of final vowels in Creek and Mikasuki, these changes now affect the final syllable of the verb stem.

For the purposes of calculating the final or penultimate syllable of the verb stem, all Muskogean languages need to make a distinction between those suffixes after the verb root that count as inside the stem ('stem-forming suffixes') and those that count as outside the stem ('non-stem-forming suffixes'). For example, in the following Creek examples, the plural suffix /-ak-/ is a stem-forming suffix, but neither the indicative /-is/ nor the second-person singular Agentive suffix /-ítʃk-/ counts as stem-forming:

- (2) Creek
 wanaj-á<h>k-is.
 tie-PLUR<PERF>-INDIC
 'They tied it (today/last night).'
- waná<h>j-ítʃk-is
 tie<PERF>-2.SING.AGENT-INDIC
 'You tied it (today/last night).'

Some suffixes may require that the preceding stem appear in a particular grade. The negative suffix /-o/ in Chickasaw, for example, must appear with the preceding verb in the glottal grade, which infixes a glottal stop in the penult of the stem.

Morphosyntactic Characteristics

All Muskogean languages show verb morphology of the agent/patient (or 'active') type, where verbs show

different types of subject and object agreement based on the semantic roles of their arguments. In Choctaw, for example, intransitive verbs fall into three types: one type uses a morphology we can refer to as Agentive agreement, another uses Patientive agreement, and a third type uses Dative agreement:

- (3) Choctaw
 Tolo:wa-li-tok.
 sing-1.SG.AGENT-PAST
 'I sang.'
 Sa-lakʃa-tok.
 1.SG.PATIENT-sweat-PAST
 'I sweated.'
 Am-ihaksi-tok.
 1.SG.DATIVE-forget-PAST
 'I forgot.'

The Dative agreement type is primarily used for the subjects of some verbs of cognition and emotion in Choctaw. In addition to these types, there is also a distinct set of Agentive prefixes with negative subjects.

Although there are clear semantic generalizations about the agreement type, the system is partially lexicalized, and verbs with similar semantics may take different types of agreement:

- (4) Choctaw
 Sa-yimmi-tok.
 1.SG.PATIENT-believe-PAST
 'I believed.'
 Anokfilli-li-tok.
 think-1.SG.AGENT-PAST
 'I thought.'

There are other areas of irregularity in the system as well, such as the fact that quantificational verbs (e.g., 'be all') take Agentive agreement in all the Muskogean languages.

Despite the existence of active verb agreement morphology, overt noun phrases are case-marked on a nominative-accusative basis, and all subjects receive nominative case, regardless of what type of agreement they trigger.

Consider the following Choctaw examples:

- (5) Choctaw
 An-ako:ʃ John(-ā)
 I-CON.NOM John(-ACC)
 ahpali-li-tok.
 kiss-1.SG.AGENT-PAST
 'I kissed John.'
- An-ako:ʃ John(-ā)
 I-CON.NOM John(-ACC)
 sa-nokʃopa-tok
 1.SG.PATIENT-fear-PAST
 'I was afraid of John.'

In these examples, both subjects receive obligatory nominative case, though one triggers Agentive

agreement and the other triggers Patientive agreement. Accusative case marking is optional in Choctaw.

Although the division among different agreement types is complex, the morphology itself is fairly regular in Western Muskogean, Mikasuki, and Creek. Mikasuki and Creek have suffixed forms of the Agentive agreement markers and prefixed forms of the Patient and Dative agreement markers. In Western Muskogean, all the agreement markers are prefixes, with the single exception of the first-person singular Agentive suffix /-li/. Though verbs vary in which type of agreement markers they use, the placement of these agreement markers is consistent across these languages.

In Alabama and Koasati, however, the situation is much more complex, with prefixed, suffixed, and infix agreement markers. Consider, for example, the affirmative conjugation of the verb /ó:tin/ 'to gather' in Koasati:

- (6) Koasati
 ó:ti-1 'I gather'
 ó<s>ti 'You gather'
 ó:t 'She/he/they gather'
 ó<l>t 'We gather'
 ó<has>t 'You (pl.) gather.'

Kimball's grammar of Koasati identifies 10 distinct conjugation subclasses based on the identity and position of the agreement morphology, with membership in one or another subclasses reflecting an intersection of lexical specification and verbal semantics.

Muskogean verbs may have a large number of both prefixes and suffixes. The prefixes mark subject, object, dative, and negative agreement; direction, instrument, applicative, reflexive, reciprocal, and location. A Muskogean verb may be followed by a large number of suffixes showing valence, causation, tense, negation, modality, adverbial modification, mood, and evidentiality/illocutionary force. Consider the following Chickasaw and Choctaw examples, which display some of the possibilities for suffixes.

- (7) Chickasaw
 Ak-tʃi-hiɬa- <ʔ>tʃ-o-ki-tok-aʔni.
 1.SG.NEG.-2.SG.PATIENT-dance-
 <GLOTTAL.GRADE>CAUSE-NEG-NEG-PAST-EVIDENTIAL
 'I must not have made you dance.'
- (8) Choctaw
 Hatʃik-im-asíɬɬ-ok-ɪʃa-k-akili-h -ô,
 2.PL.NEG-3.DATIVE-ask-NEG-YET-COMP-EMPH-TENSE-
 PARTIC.DIFF.SUBJ
 hatʃi-ki-yat ithāna-h-o:ki.
 2.PL.DATIVE-father-NOM know.DUR-TENSE-TRUE
 'Even when you (pl.) have not yet asked him,
 your (pl.) father knows.'

Word Order Properties

Muskogean languages are rather consistently head-final languages: verbs are final in the clause, nouns are final in the noun phrase, and the languages have postpositions (though these may be a subtype of relational noun rather than a distinct word class). Consider the following example from Creek:

- (9) Creek
 Má tʃokó ó:fa-n apiswa-t
 that house in-ACC meat-NOM
 ó:tʃ-i:-t ó:m-i:-s.
 exist-DUR-SAME.SUBJ be:STAT.PERFECT-DUR-INDIC
 'There is meat in that house.'

All Muskogean languages have switch-reference morphology, in which the complementizer of an embedded clause indicates whether the subject of the embedded clause is the same as the subject of the clause containing the embedded clause.

The following Choctaw example (10) shows that same-subject marking is obligatory even in contexts where there is no possible ambiguity. The Koasati example (11) shows a nice mix of same-subject and different-subject markers from a natural discourse.

- (10) Choctaw
 Ka:h sa-banna-ha:toko:-ʃ,
 car 1.SG.PATIENT-want-BECAUSE.SAME.SUBJ
 iskali' ittahobli-li-tok.
 money save-1.SG.AGENT-PAST
 'Because I wanted a car, I saved money.'
 Ka:h banna-ha:toko~,
 car want-BECAUSE.DIFF.SUBJ
 iskali' ittahobli-li-tok.
 money save-1.SG.AGENT-PAST
 'Because he wanted a car, I saved money.'
- (11) Koasati
 Já:li mók itʃo:fi-k tʃokkó:li-n
 here also uncle-NOM sit:SG-DIFF.SUBJ
 kó:si-k tʃokkó:li-n
 aunt-NOM sit:SING-DIFF.SUBJ
 i<h>l-ok
 arrive<CLAUSE.SEQUENCE>-SAME.SUBJ.FOC
 ittim-mánka-to- ~.
 RECIP.DATIVE-tell-PASTIII-PHRASE.TERMINAL
 'Here also his uncle dwelt, and his aunt
 dwelt, and he came here, and they spoke
 to each other.'

The Koasati example also shows that in some cases partial identity is enough to trigger same-subject marking. Note that in this example the verb 'arrive' is marked same-subject with respect to 'speak to each other,' though the subject of the first verb is 'he' and the subject of the second verb is 'they' – a group consisting of the subject of 'arrive' plus the previously mentioned uncle and aunt. Different Muskogean

languages (and perhaps different speakers of these languages) seem to follow slightly different principles in deciding whether partial-identity of subject is sufficient for use of the same-subject markers.

Conclusion

All the Muskogean languages are endangered; Mississippi Choctaw and Mikasuki appear to be the only languages that are currently being acquired by more than a few children. And even with these languages, there are indications that the percentage of children in the communities who acquire the language is declining.

The past two decades have seen many tribes recognizing the danger of language loss and initiating language preservation efforts. There have also been major strides in the documentation of the Muskogean languages, with the publication of grammars, dictionaries, and text collections. In the coming decades, the effort to document and preserve these languages will continue to be an issue of great urgency.

Bibliography

Broadwell G A (1993). 'Subtractive morphology in Southern Muskogean.' *International Journal of American Linguistics* 59, 416–429.

- Broadwell G A (2005). *A Choctaw reference grammar*. Lincoln, NE: University of Nebraska Press.
- Davies W (1986). *Choctaw verb agreement and universal grammar*. Dordrecht: Reidel.
- Haas M R (1941). 'The classification of the Muskogean languages.' In Spier L, Hallowell A I & Newman S S (eds.) *Language, culture and personality: essays in memory of Edward Sapir*. Menasha, WI: Banta Publishing. 41–56.
- Haas M R (1947). 'The development of Proto-Muskogean *k^w.' *International Journal of American Linguistics* 13, 35–37.
- Kimball G (1987). 'A grammatical sketch of Apalachee.' *International Journal of American Linguistics* 53, 136–174.
- Kimball G (1991). *Koasati grammar*. Lincoln, NE: University of Nebraska Press.
- Martin J & Mauldin M (2000). *A dictionary of Creek/Muskogee, with notes on the Florida and Oklahoma Seminole dialects of Creek*. Lincoln, NE: University of Nebraska Press.
- Munro P (1993). 'The Muskogean II Prefixes and their implications for classification.' *International Journal of American Linguistics* 59, 374–404.
- Munro P & Gordon L (1982). 'Syntactic relations in Western Muskogean: a typological perspective.' *Language* 58, 81–115.
- Munro P & Willmond C (1994). *Chickasaw: an analytical dictionary*. Norman, OK: University of Oklahoma.

N

Na-Dene Languages

J Leer

© 2006 Elsevier Ltd. All rights reserved.

The term 'Na-Dene' was coined by Sapir (1915) in the early twentieth century to reify his proposal of a genetic affiliation between the Athabaskan language family and two northwest coast Native American languages: Tlingit and Haida. Since then, the nearly extinct Eyak language has been shown to be closely related to Athabaskan, and the inclusion of Haida in Na-Dene has come to be regarded as unprovable by most specialists in the field. The alternative hypothesis is that the similarities between Haida and Athabaskan–Eyak–Tlingit are the result of prolonged areal contact.

Language Classification and Distribution

Athabaskan–Eyak–Tlingit (AET) consists of the Athabaskan language family plus Eyak and Tlingit, contiguous languages of the northern Northwest coast. Athabaskan is located in three main enclaves: northern Athabaskan, Pacific Coast Athabaskan, and Southern Athabaskan. Most researchers agree that the last two are the result of prehistoric migrations southward along opposite flanks of the Rocky Mountains, and that the Pacific Coast migration was earlier than the Southern. Subclassification, especially in Northern Athabaskan, is rendered difficult by the fact that these languages tend to share features with their neighbors, so that it is often difficult to distinguish the inherited from the borrowed.

The Northern Athabaskan languages are spoken in the interior of Alaska and northern Canada. Neighboring the Eskimo of southcoastal Alaska are the Ahtna, Tanaina, and Ingalik; proceeding upriver Holikachuk, Koyukon, Upper Kuskokwim, Tanana, Tanacross, and Upper Tanana are encountered. The latter, together with the Han and Kutchin (or Loucheux), straddle the border between Alaska and Canada. Proceeding southward along the Cordillera, Northern and Southern Tutchone, Tagish, Tahltan,

Tsetsaut, Kaska, Sekani–Beaver, Babine, Carrier, and Chilcotin are found. South of the Beaver on the plains live the Sarcee. Continuing in an arc through the Arctic drainage area, are the Chipewyan, Dogrib, and Slavey–Mountain–Bearlake–Hare, whose languages constitute a dialect complex.

Near the mouth of the Columbia River once lived small bands collectively called the Kwalhioqua-Tlatskanai. They appear linguistically closest related to the Pacific Coast Athabaskan of southern Oregon and northern California. Oregon Athabaskan consists of languages spoken in the interior (Upper Umpqua, Galice–Applegate) and a coastal dialect chain (Coquille, Euchre Creek, Tututni, Chasta Costa, Chetco, and Tolowa). The largest California Athabaskan language is Hupa; south of this is Mattole–Bear River and the Sinkyone–Nongatl–Lassik–Wailaki–Kato dialect complex.

The Southern Athabaskan enclave is located in the southwestern United States. The largest and best-known tribe is the Navajo, who reside mainly in Arizona and New Mexico. Most of the Apache languages are geographically and linguistically close to Navajo: Western Apache, Chiricahua–Mescalero, Jicarilla, and Lipan. The most divergent Southern Athabaskan language is Plains Apache (or Kiowa–Apache), spoken in Oklahoma.

Typological Features

The most compelling evidence for AET as a genetic grouping is the profound congruity in the makeup of the verb, as seen in **Table 1** (the lines emphasize major displacements). The AET languages have a 'templatic' structure, such that each element assumes a predetermined position relative to the others. Of particular interest are the 'classifiers,' fused combinations of what were historically separate prefixes: **l-* (which forms causatives) and **də-* (which forms passives, reflexives, etc.). An ancient irregularity is the absorption into the classifier of the stative prefix **ŋi-* which combines with *(*d*)ə- to yield *(*d*) *i-* in Athabaskan–Eyak and has become a component of

Table 1 Comparison of the verb template in the Athabaskan–Eyak–Tlingit (AET) languages

| <i>Proto-AET</i> | <i>Pre-Proto-Athabaskan</i> | <i>Eyak</i> | <i>Tlingit</i> |
|---|---|---|---|
| ⁸ Proclitic | ⁹ Disjunct prefix | ¹⁰ Proclitic | ⁸ Proclitic ⁷ Pluralizer |
| ⁷ Incorporated alienable N | ⁸ Incorporated N ⁷ Pronominal prefix | ⁹ Pronominal prefix, (u)ʔ- ⁸ Future quʔ-... ⁷ (i)lih 'mind' | ⁶ Incorporated alienable N (including pron. pref.) |
| ⁶ Pluralizer *qʔ- ⁵ Incorporated inalienable N | Pluralizer *qʔ- Lexical/derivational/classificatory prefix ^{5c} *qu-, *yʔ- ^{5d} *u- ← ----- ^{5c} *tʔ- ^{5b} *dʔ- ^{5a} *nʔ-, *sʔ-, ... | Pluralizer qʔ- ← ----- Lexical/derivational/classificatory prefix ^{5c} Gʔ-, gu- ^{5d} xʔ- ^{5c} qi-ʔ-, ləxʔ-, ... ^{5b} dʔ-, yʔ- ^{5a} lʔ- | ⁵ Incorporated inalienable N ^{5c} x'e-, ʒi-, ... ^{5b} yā- ^{5a} ka- |
| ⁴ Aspect-mode prefix | Aspect-mode prefix | Aspect-mode prefix | ⁴ Aspect-mode prefix |
| ^{4b} *ʔ- ^{4a} *Gʔ-, *s(ʔ)- | ^{4b} *ʔ-, *s- ^{4a} *Gʔ-, *nʔ-, *sʔ-, *Gu- | (ʔ), ʔ-ʔ-, Gʔ-, s(ʔ)- | ^{4d} ga- ^{4c} u- ^{4b} Ga-, na- ^{4a} Ga-, yū- |
| ³ Subject prefix *xʔ-, *ŋi-, ... | Subject prefix *xʔ-/sʔ-, *ŋʔ-, *xʔʔ-, ... | Subject prefix xʔʔ-, yi-, ləxʔ- | ³ Pluralizer daGa- |
| ² Stative prefix *ŋi- ← ----- | Stative prefix *ŋʔ- | Stative prefix yi-, s(ʔ)- | ² Subject prefix xʔa-, i-, yīi-, du- |
| ¹ Valence prefix ^{1b} *t- ^{1a} *dʔ- | Classifier *t- *(d)ʔ- | Classifier t- (d)ʔ- | ¹ Classifier ^{1c} t-, s-, ʒ- |
| ⁰ ROOT | ROOT | ROOT | ^{1b} D-component ^{1a} I-component ROOT |
| ⁻¹ Derivational-aspectual suffix *-g, *-x, ... | Derivational-aspectual suffix *-g, *-x, *-d, *-x | Derivational-aspectual suffix -g, -x | ⁻¹ Derivational-aspectual suffix -g, -x, -ʒ, -d, -xʔ, ... |
| ⁻² Aspect-mode suffix *, *-ŋ, *-t | Aspect-mode suffix *, *-ŋ, *-t | Aspect-mode suffix , -t, -kʔ | ⁻² Aspect-mode suffix ^{-2a} , -, -y, -n ^{-2b} ʒ, ... |
| ⁻³ Enclitic *-G | Negative enclitic *-he | Negative suffix -G | ⁻³ Epimode suffix -iʔ, -iʔn, -(iʔ)G |

the classifier in Tlingit. Verb stem variation also shows ancient similarities, including lengthened stem forms and in Athabaskan and Tlingit, shortening of the stem vowel before consonant suffixes. Haida verb structure exhibits some general similarities with that of AET; for example, the object pronouns precede the subject pronouns.

Other less specific typological traits are also common to Haida and AET; in particular, head-final syntactic structure (postpositions follow nouns; verbs generally come last in the sentence) and the lack or paucity of labial stops and fricatives. However, these traits are shared also with Aleut, a language that is incontrovertibly non-Na-Dene. It has been suggested that these languages, particularly Haida, Eyak, and Aleut, were part of an ancient northern northwest coast language area. If this hypothesis is borne out by archeological evidence, it may put to rest the controversy about the relationship between AET and

Haida: Na-Dene will prove to be an areal grouping rather than a genetic one.

Bibliography

- Hymes D (1956). 'Na-Dene and the positional analysis of categories.' *Am A* 58, 624–638.
- Krauss ME (1973). 'Na-Dene.' In Sebeok TA (ed.) *Linguistics in North America*, Current Trends in Linguistics, vol. 10. The Hague: Mouton.
- Leer J (1991). 'Evidence for a northern northwest coast language area: promiscuous number marking and periphrastic possessive constructions in Haida, Eyak, and Aleut.' *IJAL* 57, 158–193.
- Levine R (1979). 'Haida and Na-Dene: a new look at the evidence.' *IJAL* 45, 157–170.
- Pinnow H-J (1968). 'Genetic relationship versus borrowing in Na-Dene.' *IJAL* 34, 204–211.
- Sapir F (1915). 'The Na-Dene languages: a preliminary report.' *AmA* 17, 534–558.

Nahuatl

U Canger, University of Copenhagen, Copenhagen, Denmark

© 2006 Elsevier Ltd. All rights reserved.

Nahuatl (Náhuatl) (by some written as Nahua, Nauatl, or Nawat) is also known as Aztec and Mexicano.

Nahuatl forms with Cora and Huichol a branch of the Uto-Aztecan family and is the southernmost language of that family. Today, Nahuatl is spoken in enclaves in 10 Mexican states, from Durango in the west to Tabasco in the east. One variety of Nahuatl, Pipil, by some considered a distinct language, is spoken in El Salvador. Some 10 dialect areas have been recognized; they are different enough that mutual intelligibility is problematic. Of the 500 000 to 600 000 speakers only few are monolinguals, and the language is rapidly losing ground to Spanish.

Speakers of Nahuatl entered Meso-America some 1000 to 1500 years ago from today's northwest Mexico. The best known group, the Aztecs, settled in the Valley of Mexico, where their traditional history goes back to 1300 A.D. It was the renown of their empire that drew the Spaniards to Tenochtitlan, the capital of the Aztec empire and today's Mexico City. As a result and due to the function of Nahuatl as a lingua franca in Meso-America, Nahuatl is the most thoroughly studied and best documented of the languages in the Americas. The earliest known grammar, by Andrés de Olmos, is from 1547; by 1645 another four had been published. Olmos's grammar and Horacio Carochi's from 1645 are exceptionally accurate descriptions of the language. Alonso de Molina's dictionary of Spanish-Nahuatl (about 17 400 entries) and Nahuatl-Spanish (about 23 600 entries) from 1571 has still not been superseded. The Spanish formed the primary target group of the Nahuatl grammars and dictionaries, since they needed to speak Nahuatl in their efforts to educate and evangelize. Faced with the multitude of languages in New Spain, they even argued that Nahuatl should be the official auxiliary language, and in 1570 Philip II declared Nahuatl the official language of New Spain's Indians. As a consequence, Nahuatl was used officially in all contexts; an abundance of letters, complaints, testaments, land

deeds, and the like, written in Nahuatl, are found in university libraries in the United States and Europe and in local archives in Mexico, where new ones emerge every year. Nahuatl is thus richly documented over a period of nearly 500 years.

Under the influence of Meso-American languages, Nahuatl has shifted away from some features characteristic of Uto-Aztecan, and the last 500 years of Spanish influence is clearly evidenced. Whereas proto-Uto-Aztecan is reconstructed as verb-final, Nahuatl has verb-first word order, modified-modifier order such as possessed-possessor order, and preposed relational nouns, some of which also function as postpositions. Due to Spanish influence, former relational nouns now function as prepositions. Nahuatl basically is a polysynthetic language with person of subject and object marked on the predicate; incorporation of object and adverbials is widespread, although rare in some dialects, probably due to Spanish influence. The verb is central in Nahuatl both syntactically and in the root corpus; adjectives are derived from verb roots; and derivation is richly developed.

Phonologically, Nahuatl is uncomplicated with four vowels, short and long; 15 consonants, no voiced stops; and currently a fixed stress on the penultimate syllable

Bibliography

- Canger U (1980). *Five studies inspired by Nahuatl verbs in-oa*. Travaux du cercle linguistique de Copenhague, no. 19. Copenhagen.
- Canger U (1988). 'Nahuatl dialectology: a survey and some suggestions.' *IJAL* 54, 28–73.
- Carochi S J H ([1645] 2001). *Grammar of the Mexican language, with an explanation of its adverbs (1645)*. Lockhart J (trans. and ed.). Palo Alto, CA: Stanford University Press.
- Launey M (1992). *Introducción a la lengua y a la literatura náhuatl*. México: UNAM.
- Launey M (1994). *Une grammaire omniprédicative. Essai sur la morphosyntaxe du nahuatl classique*. Paris: CNRS Editions.
- Launey M (2002). 'Compound nouns vs. incorporation in classical Nahuatl.' In Evans N & Sasse H-J (eds.) *Studia typologica: problems of polysynthesis*. Berlin: Akademie Verlag. 113–134.

Native American Languages

E P Hamp, University of Chicago, Chicago, IL, USA

© 1994 Elsevier Ltd. All rights reserved.

The Western Hemisphere embraces about one-half the diversity of the linguistic world. Most of the populations have been always thinly settled or small, although before the coming of Europeans, Meso-America and the Andes boasted societies of density and complexity comparable to those of the Old World. Since the history of the period after about 1600 has been one of frequent linguistic retreat in the face of European intrusion, an appreciable portion of the earlier languages, dialects, and even families has vanished from the opportunity for knowledge and our dossier, with the result that a simple catalogue shows a large gray area among the moribund, the extinct, the uncertain, and the unknowable. The total number of linguistic stocks which are believed by informed and responsible linguistic scholars to survive in the New World unreduced by late twentieth-century methods of genetic comparison reaches about six score families and six dozen isolated singletons; it is realistic to hope for a modest reduction in these numbers by the theory and method accepted in the 1990s, but it would be either foolish or revolutionary for a linguist to expect a division of these totals by anything larger than the smallest numbers.

The mere recitation of theories proposed to account for the geographic and social distributions in which these languages and stocks are found would greatly exceed the space available here. Suffice it to say that there exist in the literature (cited and referred to) theories that are better than speculative relating to the past couple of millennia and also approaching the past tens of millennia and which argue for the presence of these languages both on internal principled linguistic criteria (often fashioned on Old World models as well as modern structural and typological arguments) and on correlations with archaeology, ethnography, folklore, and cultural theory. In North America therefore the study of these languages has been closely linked with the field of anthropology, and in turn has had a strong influence on the development of the course of linguistic theory in the twentieth century.

The Size of the Problem

These languages are not only too numerous to permit a useful itemization here or even in more specialized articles, but they exemplify typologically almost all linguistic features which have been identified in the Old World, with rare exceptions, e.g., the Khoisan

clicks, the intercalated affixes of Semitic, the marathon domains of a finite verb in Mongolian discourses, or perhaps the feature/segment ratio of North Caucasus phonologies. The New World presents in addition the immense and internally complex words of Eskimo; the polysynthetic word structure (see Rood 1992) emphasized in the writings of Sapir (especially his book *Language* 1921) and found in northern North America, in Totonac, and in much of South America; and the Oto-Manguean Mazatec and Chinantec complex tone glides which englobe tonal inflexional suffixes and make highly explicit whistle speech possible. The polysynthesis has important bearing on grammatical theory by erasing or altering the boundary between morphology and syntax, a boundary which is difficult to specify or justify in diachronic linguistics; it seems rather that morphology is itself a characterizing feature of certain languages and stocks. Oto-Manguean tones, like Khoisan clicks, raise an interesting question of rarity; such features can be dismissed as atypical because rare (Rood 1992: 112), and if so one might class them as marginal to the 'universal' and 'natural' capacities of human language. Yet these may also be viewed as precious relics left behind by history to attest to and suggest further the possible range of natural language.

The explicit sentence and discourse morphologies of the Americas (quotatives, deixis, obviatives, switch reference, noun tenses), in addition to verbal morphologies (with special Northwest forms for embedded noun complements) far exceeding the categorical scope of Homeric and classical Greek, distinguish these languages as independent witnesses to the systematic refinement of the human mind, and make the rough clod of many a European word appear as a shapeless mass and intruder in a considered sentence.

The big lesson from all of this is that human language as it is known, while differing enormously from place to place and society to society in detail, seems to repeat a somewhat closed inventory of features which students of culture would define as adequate for any society of such creatures as man. There is not space here for even an outline of all the detail, and the appended bibliography should be regarded as an extension and key to indefinite elaboration of this article. This arrangement also rests on and emphasizes the conviction that science must be cumulative. The contribution of the New World has been twofold: the confirmation of Old World evidence (even down to the testimony of Egyptian and Mesoamerican inventions of writing through a progression of pictographic to logographic to syllabic to alphabetic representation) for the Jeffersonian percept of the unity

of man; and the supply of a vast wealth of detail for analysis and formulation of the diversity and flexibility possible within this unity. In the statement of the exact mechanics of the first and in the exploration of the second there remains much and urgent unfinished business for future scholarship.

Scope of this Article

This article can provide only a coarse discursive chart. The stocks are surveyed genetically because in this way one can be assured that the review is (in principle) coexhaustive. Areal analysis is valid, interesting, and important (cf. Campbell 1992, and the work of J. Sherzer for North America), but such study, classic in the Balkans and the Caucasus, is grossly deficient as yet in the Americas, and such an aspect as a basis does not assure total coverage of the subject in a limited space. The recitation of typology is necessarily eclectic; and the reader is referred to the bibliography which will also reflect the evolving views of linguistic theory.

The coverage here of the various stocks varies. Some have been voluminously studied, and the interested reader can get information in appropriate detail elsewhere; the separate articles in this Encyclopedia will supply what is not repeated here, and the bibliography is meant to lead successively to the fuller literature – that is the main reason for some titles. Greater detail (in brief) is given for some stocks not represented by separate articles. For reasons which will become clear below, the stocks of South America are not treated with comparable completeness and fullness. Close geographic location is not given here; space forbids the necessary verbosity, and sufficiently detailed maps are voluminous and hard to read. Ethnographic maps may be used, but one must remember that tribe or social unit is not language. No inventory of the hundreds of separate languages is approached here; many exact and complete listings are included in the bibliography.

The Identification of Linguistic Stocks

The classifications which have been made for these languages naturally reflect the development of orderly principles which has unfolded in the history of linguistics during the nineteenth and twentieth centuries, a time span which happens to coincide with the growth of serious knowledge and census-taking of languages and societies in aboriginal North America since the expedition of Lewis and Clark dispatched by Thomas Jefferson in 1804–06, the first systematic inventory and collection of languages and data. The nineteenth century was largely occupied with simply collecting,

discerning, and sorting the varieties and with describing grammars and compiling dictionaries and glossaries only for languages of white man's closest contact. The collection and census activities have continued with increasing refinement and changing emphasis down to the present; in the late twentieth century a new interest has grown in the presentation of exact and analyzed texts. While several nineteenth-century grammars are valuable as working tools to this day, it is the twentieth-century North and Middle American production of descriptive Native American grammars and topical articles or grammatical segments that has drawn widespread attention and interest in the world of linguistics; indeed many of these have been models and proving grounds for theory making.

The earlier classificatory work in the nineteenth century had a minimal theoretical basis and relied heavily on vocabulary inspection, not through naïveté and neglect of reflection but from a misguided report that all American languages shared the same grammatical peculiarities. Thus the work of Duponceau (1819), Gallatin (1836), and Trumbull (1876) was scarcely touched by the emerging comparative method of Europe. Though Daniel G. Brinton presented a careful inspectional survey in 1891 in which he put the inventories of families at about 80 for each of North and South America and from which all later South American classifications in some measure derive, he did not, unfortunately, have access to the contemporary classification of Powell (1891), a geologist and army man aided by the ornithologist. H. W. Henshaw in most of his decisions, an inspectional lexical classification into 58 families none of which has since needed to be dismantled but a classification without analysis and nearly without any more linguistic theory than common sense.

Then, in the twentieth century, more abstract modes of comparison have been attempted, progressively exploiting more complete and detailed descriptive grammatical knowledge. Notable in this reduction of claimed familial stocks is the proposal of Sapir (1929) with six phyla for all of North America: Eskimo-Aleut, Na-Dene, Algonquian-Wakashan, Aztec-Tanoan, Penutian, Hokan-Siouan. To arrive at these Sapir drew on the comparative method, which he understood well, extended by single abstract paired features (e.g., in Penutian), by typology (in Na-Dene and Hokan), by intuition from his vast firsthand fieldwork contact (Algonquian-Wakashan), by geography (Hokan-Siouan, Athabaskan), by ignoring remainders (e.g., Keresan, Yukian). Further attempts were made to manipulate these, e.g., to reattach Algonquian and the Southeast US (including Muskogean) languages; the latter pairing unfortunately failed to adduce multiples of the claimed phonetic

correspondence sets. Hokan comparisons were flawed by unmotivated deletions in the phonological strings within lexical comparanda. Abandoning these reductionist attempts, since the mid-1960s specialists have favored less speculative groupings: Voegelin and Voegelin published in 1966 a map based on a 1964 conference which is reflected in Bright (1974), and which returned to about 15 stocks (repeated in Sebeok 1973) shown in Bright as 7 phyla and 9 individual families (including isolates), the phyla closely matching Sapir's six (for a later version see Voegelin and Voegelin 1977). A further conference in 1976 produced Campbell and Mithun (1979), which is reflected in Rood (1992), where about 60 families are recognized for North America, of which about 37 lie west of the Rockies. These families are almost all formulated on criteria invoking the comparative method; Sapir's phyla, and their revisions, rest, per contra, on the hope that ultimately the comparative method, in some refined version, may confirm those conjectures. In the 1990s, if the comparative method will not apply, i.e., show a unique result, one cannot assert a genetic relation.

Because the North American stocks, especially Uto-Aztecan, form a continuum into Middle America, they are here considered together. Therefore, added to the ± 60 of North American are the 3 Middle American families (Oto-Manguean, Mayan-Mixe-Zoquean, and Totonacan) and half-dozen isolates (Ta'araskan, Huave, extinct Cuitlatec, Xinca, Tequistlateco, and Jicaque).

For South America the pure survey work is not yet finished; yet languages are dying out, and the situation is urgent and grave. This inchoate stage of scholarship is reflected in authoritative accounts: Manelis Klein (1992) mentions 32 families and 74 singles, while Kaufman (1990) meticulously discusses about 60 families (agreeing closely with Loukotka 1968) and 60 singles. One cannot further compress their compact arguments, but the complexity and sheer bulk of the problem, and the need to distinguish knowledge from simplistic ignorance, is apparent.

In the late 1980s there appeared (Greenberg 1987) a proposal, adumbrated in earlier publications, to reduce the stocks of the New World (and even Old World phyla) to but three: Eskimo-Aleut, Na-Dene (including the disputed isolate Haida), and Amerind (all the rest). This is really not a proposal comparable to accepted work in genetic and diachronic linguistics; the earlier long-range comparisons of Swadesh were different in detail and in principle, being rather a bold and radical extension of Sapir's reasoning. No one who knows the data and the literature doubts the unity of Eskimo-Aleut, and of Athabaskan-Eyak with Tlingit (especially with the Tongass dialect

prosodics in relation to AE tones, which effectively eliminates any possible direct relation to Sino-Tibetan). Greenberg's proposal is therefore simply the insistence, in a naïve version, of Duponceau, et al., that all languages in this chosen area are related, and are to be called 'Amerind'; the criterion invoked is just subjective partial phonetically uncritical similarity with no constraints of the degree demanded, e.g., by the comparative method.

The Stocks of North and Middle America

Eskimo-Aleut

This compact family occupying the Arctic coast and adjacent islands from easternmost Siberia eastward to include Greenland shows a considerable time-depth of divergence between its two main members. It includes a recognized national language in Greenland, official regional languages in Canada (with its own syllabic script) and Alaska, two very different Eskimo languages in Alaska, and a distinct regional variety in Siberia. It is likely though not yet shown conclusively that this family is related as next of kin to the Luoravetlan languages of eastern Siberia and a relation has further been argued, on the basis of carefully sifted data, with Yukagir and Uralic. Considerable scholarship exists correlating this linguistic family with findings of archaeology. It is reasonable that this family arrived in the New World having crossed the Bering Strait separately.

Na-Dene

The Athabaskan (or Athapascan) family is known in great detail, and reconstructions of its forms and grammar are comparable to those of Indo-European and Romance, which is not the case with Eskimo and Aleut. The Navajo-Apache branch in the Southwest US and the Pacific Coast branch in Oregon (extinct) and in northwesternmost California may be viewed as outliers to the Northern branch in interior Alaska and northwestern Canada; Sapir constructed a classic proof of the northern origin of the Navajo. Eyak, now extinct, was described in ample and modern fashion by salvage work with the last three speakers, and proves invaluable as a sister branch to all of Athabaskan. Tlingit still presents problems of exact comparison analogous to those found in Altaic.

Algonquian

Algonquian is the most completely known family of the New World, with extinct languages of Eastern Algonquian on the Atlantic seaboard shedding light on facts from the time of first British and French contact yet raising philological problems from those

early records. Analogues to the reconstruction of Indo-European have been practiced on Algonquian, including the beginnings of sophisticated dialectology for Cree and Ojibwa and the reconstruction of a Heimat in the southern Ontario Great Lakes region. Much to most linguists' surprise Sapir's conjectural hypothesis that diminutive Wiyot (now extinct) and Yurok on the north California coast are each related to all of Algonquian proved correct; this three-branch family may be called 'Algic.' Ritwan does not name a genetic unit. The internal subgrouping of non-Eastern Algonquian is not at all clear or agreed.

Iroquoian

Also very well known from colonial times, Iroquoian has been admirably studied, and falls neatly into a Northern branch and the southern Cherokee. There are strong suspicions that Iroquoian is related to Siouan and Caddoan, and perhaps even to Yuchi, but no firm proof has yet been offered. The fact that Sequoya, a Cherokee, invented a syllabic writing does not help philologically, but adds to the dossier of graphic invention.

Siouan and Caddoan

These groups share the Great Plains north of Texas, and Siouan also occupies, or formerly occupied, eastern Wisconsin and portions of the southeast US. Catawba, in the Carolinas, certainly forms a distant separate branch, but the subgrouping of the other Siouan branches is still not sure: the southeastern, or Ohio Valley, branch comprised Ofo, Biloxi, and Tutelo; the Missouri River branch embraces Crow and Hidatsa; and the Mississippi Valley branch all the other languages, including Dakota, Chiwere, and Quapaw. Siouan, on which the late twentieth century has supplied great gains over earlier scholarship, is notable for its suppletion between singular and plural.

Caddoan

Neglected until the late twentieth century, Caddoan comprises Pawnee, Arikara, Caddo, and Wichita. The last of these presents one of the lowest phoneme counts in the world (for a convenient example of polysynthesis by a specialist in Wichita see Rood 1992: 113).

Muskogean

Although the subgrouping of this family, which was located entirely within the southeastern US before displacements to Oklahoma by the European immigrants, is not yet decided, the relation of Choctaw-Chickasaw, Alabama-Koasati, Hitchiti-Mikasuki, and Creek (including Seminole Creek) is assured by exact and sophisticated reconstruction. Whether a

'Gulf' unity may be approached by the attachment of Natchez has not yet been demonstrated. Even less convincing is the genetic addition of the isolates Tunica, Chitimacha, Atakapa, and Yuchi of this south-eastern area, of the extinct but philologically documented Timucua in Florida, and of Yukian (including Wappo) in coastal California, all of which Sapir attached to his 'Siouan.' In fact the case for Yuchi looks slightly more hopeful than the others.

Keresan

Keresan comes from of the pueblos of New Mexico; it was also included by Sapir in his 'Siouan' on evidence that would not convince many in the early 1990s.

Salish or Salishan

This is a compact family of two dozen languages extending along the Pacific coast northward from Oregon through the southern half of British Columbia and inland into the Rocky Mountains; these languages, typologically similar, show about the degree of divergence seen in Romance, and have been admirably explored and documented during the last third of the twentieth century. Many now have grammars and dictionaries. The subgroups are: the northernmost Bella Coola in coastal British Columbia, which is a sibling to all the rest; Coast Salish, with about ten languages; the now extinct Tillamook in Oregon, which lacks labials; Tsamosan, comprising two groups, Upper Chehalis and Cowlitz, Quinault and Lower Chehalis; and Interior Salish, with a Northern branch comprising three languages including Shuswap and Lillooet, and a southern branch of four including Kalispel and Coeur d'Alene. Bella Coola permits syllables and sizeable words with no vowels; some of these languages rank next to the Caucasus in number of consonants and richness of simultaneous combinations of distinctive features. An interesting rarity found in Interior Salish is glottalized sonants. The comparative morphology of these isomorphous languages is fairly straightforward and conducive to elegant refined formulation; good progress has been made with comparative syntax.

Wakashan

Wakashan is a small but intensively and long-studied family. There are only a half-dozen languages in the Vancouver Island region, divided into a Northern or Kwakiutlan branch and a Southern or Nootkan one; these branches are not quite so different in their divergence as Indo-European branches are. An intrinsic interest in the scholarship of Wakashan is that it has attracted the attention of Boas, Sapir, and Swadesh as well as fine contemporary scholars. The phonology of

these languages includes a rich array of laryngeal and pharyngeal phonemes; much of the talk of complex Amerindian morphology, of the alleged gray area between the classes noun and verb, and of the character of Amerindian lexicon which is strange to Europeans has come from the study of Wakashan. To appreciate these claims and their truth or traditional fiction a linguist would do well to study both the earlier and ongoing scholarship on this family.

The Chimakuan family on the Olympic peninsula in the state of Washington consists of only the extinct Chemakum and the moribund Quileute. Attempts so far to connect Chimakuan and Wakashan lead perhaps more to old loans and contacts than to genetic relation. It was long attempted to unite Chimakuan, Wakashan, and Salishan under the name 'Mosan' (from the numeral '4'); certainly some features, such as the presence of the so-called 'lexical suffixes' (as verb complements), appear as encouraging comparanda, but they may well be simply Sprachbund diffusional traits. A few numerals will hint at the difficulties:

| | | | | | |
|------------------|--------------|----------------|-----------|-------------|---------------|
| '2' <i>Chim.</i> | *təʔ(a)kw(a) | <i>Nootk.</i> | ʔaχ-a | <i>Sal.</i> | ʔsa:l- |
| | | <i>Kwak.</i> | m(a)ʔ(a)+ | čū- | tVq- |
| '3' <i>Quil.</i> | qʷa'al | <i>Nootka-</i> | qakčə | | čan, |
| | | <i>Nit.</i> | | | |
| | | <i>Makah</i> | wii | | †exʷ, čʔ+V(?) |
| | | <i>Kwak.</i> | yudxʷ | | |
| '4' <i>Chim.</i> | *maʔ(a)yas | <i>Wak.</i> | *muy | | mu:s |

Penutian

Penutian is the name which has been applied to a set of languages embracing central California, much of Oregon and southern interior Washington, Tsimshian in coastal northwest British Columbia, and Zuni in western New Mexico (these according to Sapir's classification), to which have been added at times Mixe-Zoquean, Totonacan and Mayan, which here shall be kept separate. In the later twentieth century many of these, mostly in Oregon, have been extinct, but fortunately they were largely competently documented, if to a limited degree and quantity; for example, Sapir's classic work on Takelma and material still to be published on Kalapuya. Thanks to enormously copious, detailed, and accurate work of the second half of the twentieth century we are increasingly well informed on the surviving languages, so that the position can now be surveyed, if not with certitude, at least with competence and focus not possible in the first quarter of the twentieth century. California Penutian (Miwok-Costanoan, Wintun, Maiduan, Yokutsan, all with numbers of small languages, and probably Klamath-Modoc in southern Oregon) certainly makes up a family. The unity of Takelma-Kalapuya and Coos with Alsea-Siuslaw as Oregon Penutian is

less clear; the relation of Chinook and Sahaptin-Nez Perce to an Oregon Penutian unity can be rated only as probable. A Plateau Penutian with Cayuse and Molale is not very clear at all, while the relation of Tsimshian to all of these remains as yet unclarified or even undemonstrated. A relation of Zuni to California Penutian has been asserted, but there is no sign that this claim has met with significant acceptance.

The languages and branches mentioned above present a highly varied typology, and it seems scarcely possible that many valid relations could be discovered by simple inspection without searching grammatical analysis and the formulation of well motivated structural interstages. Some of the members of California Penutian show inflectional paradigms and word shapes that look startlingly like Indo-European, while to the north one encounters specimens of surface structure from another world. The problem of Penutian remains complex and diffuse.

Uto-Aztecan and Kiowa-Tanoan

The unity of the much studied and restudied widely dispersed and highly diverse Uto-Aztecan, extending from Yellowstone Park in Wyoming to the Aztec empire (including Nahuatl and Pipil in Mexico and Central America), cannot be called into question. The unity of the small compact family comprising Kiowa of the Oklahoma Plains and Tanoan (Tiwa, Tewa, and Towa), sharing the New Mexico pueblos with Zuni and Keresan, is likewise assured. But the Aztec-Tanoan relation which has been claimed is not on the same level of certitude. The political importance and vast scholarship attaching to Nahuatl places this language in a different class of cultural interest from all except a few of the other languages discussed in this article.

Hokan

Perhaps even more than Penutian, Hokan is a problematic proposed stock; the name has been applied to a large number of languages and small-to-modest sized families in the western margin of the continent from northern California south to Oaxaca; Sapir's classification included even more, notably so-called 'Coahuiltecan,' of northern Mexico and adjacent Texas, and even attempted to attach Siouan. Here one can attempt to do little more than inventorize the groups that can be discerned among these sharply divergent language structures.

The Yuman family of southern California, Arizona, and Baja California is well understood and has been admirably researched in the second half of the twentieth century; likewise the Pomo family of the coast north of San Francisco has been probingly studied, and it seems likely that these two are related. It is also

claimed that Seri, on the northwest coast of Mexico, and carefully studied, is related to these. Further possible members of uncertain kinship as yet, are found in California (all but three of these are now extinct): Karuk, Shastan, Palaihnihan (two languages), Yanan, Chimariko, all in northern California; Washo, on the Nevada border; Esselen, Salinan, Chumashan, on the southern California coast.

The allegiance of Tequistlatecan is still moot.

Isolates of southern Texas and Mexico

In southern Texas and adjacent Mexico there was a string of languages, now extinct, which used to be assigned to Hokan, but are now probably to be regarded as isolates or simple relics: Tonkawa, Coahuilteco, Karankawa, Comecrudan, Cotoname, Solano, and Aranama.

North American Isolates Resisting Affiliation

Isolates of North America which should not be overlooked are Haida on the Queen Charlotte Islands and Kutenai of the Idaho-Canada border, both well studied; and the extinct and obscure Beothuk of Newfoundland.

Mayan and Mixe-Zoquean

The large and compact Mayan family of southern Mexico and Guatemala is one of the best studied language families in the world, with much ongoing research continually reported on; it would be absurd to pretend to outline or indicate this vast terrain of activity in the present inventory. A unique feature of this New World family and of the last half or third of the twentieth century is the writing system, its decipherment, and the correlation made between the dialectology of this koinē writing and the surviving spoken languages. The decipherment of Middle American writing gives nearly two millennia of linguistic history, apart from the content of the texts.

The Mixe-Zoquean family, with approximately 90 000 speakers, seems to be relatable to the nearby Mayan family. The decipherment in 1992 of the 159 AD proto-Zoqueanda Mojarra stele, which antedates Maya inscriptions and possibly reflects Olmec culture, surprisingly has pushed back the history of New World writing.

Totonacan

Totonacan is a family of just two languages on the eastern coast of Mexico, with highly agglutinative structure; Totonac numbers 265 000 speakers, and Tepehua 8000. It has been suggested that Totonacan is related to Mayan (for documentation see McQuown 1990).

Oto-Manguean

This family, poorly investigated until around 1940 had become extremely well surveyed and basically described and outlined by around 1970; since then much descriptive detail has been filled in, although the complexity and number of these languages and dialects leave much essential work still to be done, and comparative work has been less active since around 1975. The record of work and publication is, however, formidable and inspiring. Oto-Manguean is a completely Middle American family, perhaps the oldest established one; its typology is highly original and distinctive; the languages are many, divergent, and demanding of careful analytic concentration; the scholarship has shown resourcefulness and great diligence. Oto-Manguean is the only known stock in the New World that shows a degree of divergence with a number of branches comparable to that of Indo-European. The amount of time mentioned above for the exploration of this family gives a notion of the scholarly achievement which is found here.

Except for Otopamean, which is found in central Mexico, this family is located largely in Oaxaca. The branches, some with a dozen or so languages each, are:

- a. Chiapanec-Mangue (extinct, in Chiapas, and Nicaragua and Costa Rica; two languages);
- b. Otopamean: Otomian (one-half million speakers; Mazahua and Otomi), Matlatzincan (2000 speakers; Ocuilteco and Matlatzinca), Pame (4000 or 5000 speakers; North and South), Chichimeca Jonaz (1200 bilinguals);
- c. Mixtecan: Mixtec (285 000 speakers), Cuicatec (20 000 speakers), Trique (16 000 speakers);
- d. Zapotecan: Chatino (32 000 speakers), Zapotec (many languages);
- e. Popolocan: Mazatecan (120 000 speakers, many varieties with a complex dialectology already studied), Chocho (2500 speakers), Popoloca (at least 10 000 speakers), Ixcatec (one or two speakers in 1969);
- f. Chinantecan: 60 000 speakers, a dozen varieties;
- g. Amuzgoan: 30 000 speakers in three varieties;
- h. Subtiaba-Tlapanec (the first extinct, in Nicaragua; the second 40 000 speakers in Guerrero): disputed as being Hokan.

As yet there is not a satisfactory subgrouping for these branches. Probably the most remarkable single feature of almost all these branches is the complexity of their tonal systems, which often participate in an intricate way in their inflectional systems; added to this may be a set of involved morphotonemic rules.

Tarascan

Tarascan is an isolate in Michoacán spoken by over 60 000, for which a descriptive sketch will be found in Edmonson (1984).

Remaining Isolates

Remaining isolates in Middle America which are not attached linguistically to South America are Huave, in the Isthmus of Tehuantepec, with 6000 speakers; Cuitlatec extinct in Guerrero; Xinca, which has been overrun by the Maya; and Jicaque, of Honduras, which does not belong to the Mesoamerican Sprachbund (on the Central American languages that form a transition to South America see Craig 1985).

The Stocks of South America

The problem of simply inventorizing the stocks of South America has already been raised in Sect. 3 above. There is no point in further belaboring the predicament in presenting an intelligible and coherent picture at this juncture of fact-gathering activity. Fundamentally, more field data is needed, more descriptive analysis, more basic lexica, more dialectology – all of this simply to know what is being counted. Then there must follow a redoubled effort at linguistic comparison, using areal analysis and typological criteria as heuristic controls in order to identify by elimination the probable inheritances. In that process of comparison it must be hoped for an acceleration of responsible discovery comparable to that chronicled for work in Siouan, Salishan, Penutian, some parts of Hokan, and especially Oto–Manguean.

It is estimated that in the whole of South America, of the approximately 500 languages spoken at the time of European contact about 300 survive. For a critical discussion of the present status and the task ahead, and for an explicit inventory of the results of such a criticism Kaufman (1990) is the most useful and most objective resource. At the end of the twentieth century the clearest presentation of stocks which have been discerned is an areal one, based on simple geography.

Lowland South America

This area presents the largest number of stocks which have been identified and reasonably delimited up to the early 1990s; it includes the complex region known as 'Amazonia.'

In the north and verging into Central America one finds Chibchan. A very widespread family is Arawakan, and another that is prominent is Cariban. An hypothesis also exists that Tupian is related to Cariban. Towards the Andes is the Panoan family.

An important but smaller family is Tukanoan. And the exact constitution still needs to be established for Gê (Jê), or Macro-Gê.

The Southern 'Cone' Area

This southern portion of the continent is characterized by far greater linguistic fragmentation – either by virtue of our ignorance or in terms of multifarious small residual stocks and isolates. Some families that can be named here are Araucanian, Chon, and Guaycuruan. The large, important and widespread Guaraní language, which is official in Paraguay and spoken also in surrounding territories, is particularly to be noted; and along with this compare also Tupi.

Languages of this area are still becoming known thanks to the welcome increase in the mid- to late twentieth century in descriptive work by well-prepared linguists.

The Andean, or Highland Area

As a result of events, mainly conquest, both before and after the coming of the Europeans the linguistic complexity of this area is much smaller in terms of languages than that of the other areas, but it is correspondingly greater in terms of geographic and social dialectology.

Two languages dominate attention in this area, Quechua (or Quichua) and Aymara. These two typologically form a linguistic area, along with others, and have in the past been wrongly claimed as kindred members of the same stock. It is now clear to competent scholars that Aymara, with around two million speakers, is to be classified with Kawki and Jaqaru in the Jaqi family, while the giant Quechua, with around ten million, must remain for the time being classified as an isolate.

In the earlier colonial period Puquina, now extinct, was important.

Comparatively well studied among the small surviving languages are the pair Uru and Chipaya in Bolivia.

For the remainder of an ongoing task the reader is referred to the references in the bibliography.

Bibliography

- Albó X (1979). A sociolinguistic panorama of Central and South America. In Kloss H & McConnell G D (eds.) 1974–84.
- Bergsland K (1974). Eskimo-Aleut languages. In *Encyclopaedia Britannica*, 15th edn. Chicago, IL: Encyclopaedia Britannica Inc.
- Boas F & Hamp E P (eds.) and author, (1911, 1922, 1933–38). *Handbook of American Indian Languages*.

- Bureau of Ethnology Bulletin 40. Washington, DC: Government Printing Office.
- Bright W O (1974). North American Indian languages. In *Encyclopædia Britannica*, 15th edn. Chicago, IL: Encyclopædia Britannica Inc.
- Campbell L (1992). Meso-American languages. In Bright W O (ed.) *International Encyclopedia of Linguistics*, vol. 2. Oxford: Oxford University Press.
- Campbell L & Mithun M (eds.) (1979). *The Languages of Native America: Historical and Comparative Assessment*. Austin, TX: University of Texas Press.
- Craig C G (1985). Current knowledge of Amerindian languages of Nicaragua. *IJAL* 51, 381–384.
- Derbyshire D C & Pullum G K (eds.) (1986). *Handbook of Amazonian Languages*, vol. 1. Berlin: Mouton de Gruyter.
- Edmonson M S (ed.) (1984). *Linguistics*. (Supplement vol. 2 to the last). Austin, TX: University of Texas Press.
- Greenberg J H (1987). *Language in the Americas*. Stanford, CA: Stanford University Press.
- Hoijer H *et al.* (1946). *Linguistic Structures of Native America*. New York: Viking Fund.
- Kaufman T (1974). Meso-American Indian languages. In *Encyclopædia Britannica*, 15th edn. Chicago, IL: Encyclopædia Britannica Inc.
- Kaufman T (1990). Language history in South America: What we know and how to know more. In Payne D L & Hamp E P (eds.).
- Kloss H & McConnell G D (eds.) (1974–84). *Linguistic Composition of the Nations of the World/Composition linguistique des nations du monde*, 5 vols. Québec: Presses de l'Université Laval.
- Loukotka Č (1968). *Classification of South American Indian Languages*. Los Angeles, CA: UCLA Latin American Center.
- McQuown N A (ed.) (1967). *Linguistics*, Handbook of Middle American Indians, vol. 5. Austin, TX: University of Texas Press.
- McQuown N A (1990). *Gramática de la lengua totonaca*. México: Universidad Nacional Autónoma de México.
- Manelis Klein H E (1992). South American languages. In Bright W O (ed.) *International Encyclopedia of Linguistics*, vol. 4. Oxford: Oxford University Press.
- Manelis Klein H E & Stark L R (eds.) (1985). *South American Indian Languages: Retrospect and Prospect*. Austin, TX: University of Texas Press.
- Merrifield W R & Rensch C R (1990). *Syllables, Tone, and Verb Paradigms*. SCHL 4. SIL. Arlington, TX: University of Texas.
- Payne D L & Hamp E P (eds.) (1990). *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin, TX: University of Texas Press.
- Pinnow H-J (1964). *Die nordamerikanischen Indianersprachen*. Wiesbaden: Harrassowitz.
- Powell J W (1891). Indian linguistic families of America north of Mexico. In *US Bureau of American Ethnology Annual Report*, 7. Washington, DC: Government Printing Office.
- Rensch C R (1976). *Comparative Otomanguean Phonology*. Language Science Monographs, vol. 14. Bloomington, IN: Indiana University.
- Rensch C R (1989). *An Etymological Dictionary of the Chinantec Languages*. Studies in Chinantec Languages, 1. Arlington, TX: SIL, University of Texas.
- Rood D S (1992). North American languages. In Bright W O (ed.) *International Encyclopedia of Linguistics*, vol. 3. Oxford: Oxford University Press.
- Sapir E (1929). Central and North American Languages. *Encyclopædia Britannica*, 14th edn. Chicago, IL: Encyclopædia Britannica Inc.
- Sebeok T A (ed.) (1973). *Current Trends in Linguistics*, vol. 10. The Hague: Mouton.
- Sturtevant W C (ed.) (1978–). *Handbook of North American Indians*. Washington, DC: Smithsonian Institution.
- Suárez J A (1974). South American Indian languages. *Encyclopaedia Britannica*, 15th edn. Chicago, IL: Encyclopaedia Britannica Inc.
- Tovar A & Larrucea de Tovar C (1984). *Catálogo de las lenguas de América del Sur*. new edn. Madrid: Gredos.
- Voegelin C F & Voegelin F M (1977). *Classification and Index of the World's Languages*. New York: Elsevier.
- Waterhouse V (1985). True Tequistlatcco. *IJAL* 51, 612–614.

Native Languages of North America, Variation in

S Trechter, California State University,
Chico, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

The concept of variation has framed discussions of Native North American linguistics from its beginnings. In his introduction to the *International Journal of American Linguistics* in 1917, Franz Boas summarized what direction such a journal should

take toward further understanding of the complexities within and the relationships between American indigenous languages. His outline and exhortations serve to broadly define the scope of linguistic variation for indigenous language study: for the purposes of dialect and language identification, linguists should extend their examination beyond lexical cognates and sound correspondences to an analysis of morphological variation existing in polysynthetic languages. Recognizing linguistic change induced by daily contact with

speakers of European languages, comparison across generations would be vital. Particular attention should be paid to examining the variation that occurs between different genres of speech from conversation to folktales to ritual performance in order to capture lexical variety as well as literary convention. The linguist should be aware of not only individual variation such as that associated with a poet, but also the types of variation that may be conventionalized across particular groups of people. Such a broad focus is more akin to ethnography of speaking than mere study of variation, and this focus is integral to research in Native North American languages.

Much indigenous language variation is overtly recognized as a genre or identity marker by community members, rather than the minute, subconscious indicators of gender, class, and ethnicity described in the study of more widely spoken languages. As a personal characteristic of everyday speech, variation indicates a particular position taken by a speaker with respect to his interlocutors in a context. Some indicators of a speaker's identity may be less contextually malleable than others – such as regional background. Other variations that occur may indicate personal characteristics of the speaker or addressee (gender, age, physical handicaps) or intentional affect. Interestingly, though such 'exotic' forms have often been attributed to North American languages, they are not typically categorical. They are often associated with specific genres of speech or circumstances and give rise to a variety of expressive nuances. Finally, some variation is almost completely determined by genre, such as ritualistic or shaman's language, announcements, prayers, and folktales, rather than the identity of the speaker. However, speakers' abilities to engage in certain kinds of formalistic language may also indicate something about their overall position in society and certainly their position of authority in a ritualistic context. Despite language attrition, which affects all kinds of language variation, Native North America is rich in linguistic variety and the variation exhibited.

Regional Dialects and Mutual Intelligibility

Given the large number of languages, approximately 300 north of Mexico, determining the difference between dialects and languages and their similarities is a process of constant revision. Many languages such as Yuchi, Zuni, or Takelma are apparent isolates. Others, however, are parts of extensive dialect continua, such as Ojibwe (Ojibwa) (Algonquian), Western Apache (Athabaskan), Straits Salish (Salish), or Dakota (Siouan). The variation that exists from one

community dialect to the next can be an important marker of ethnic identity, and where this ethnic distinction is particularly emphasized, some speakers may be willing to sacrifice mutual intelligibility with other dialects to maintain their distinctiveness. In such circumstances, native speakers are quite willing to criticize each other for mixing forms from another dialect or language, or for innovating too much. The languages of the pueblos in the southwest are particularly famous for such emphasis on linguistic purity. The Arizona Tewa (Kiowa-Tanoan) strongly proscribe language mixing in *kiva* speech and overtly maintain that Tewa is 'pure' despite multilingualism and continual contact with the Hopi (Uto-Aztecan) for the past 300 years. Indeed, there are almost no adoptions of vocabulary from Hopi into Tewa, but there appears to be considerable influence from Hopi phonology and grammar. Likewise, the White Mountain Apache claim their linguistic separation along a dialect continuum from Navajo to other Western Apache dialects from Tonto/Camp Verde-White Mountain-San Carlos through overt comments about mixing:

That's the way that they talk at San Carlos . . . that's not our word, that's a Navajo word . . . when he goes to Camp Verde he comes back and says things like they say things. That's wrong, we don't say them like that over here. He needs to stay in one place. (Greenfield, 1999: 375)

In other dialect continua, such as among the Creek (Muskogee) (Muskogean) of the southeastern United States, Dakota (Siouan) of the Great Plains of the United States and Canada, or the Ojibwe (Algonquian) in the Canadian regions of the Great Lakes, strong phonological or lexical shibboleths are noted or occasionally mocked when a speaker moves from one community to another, but there is more tolerance for mixing. Attitudes towards archaism, purity, mixing, and new languages are as varied as the dialects themselves.

Assessing Variation

Native speaker judgments concerning dialect mixing are of some use in establishing what should be counted as variation between dialects; yet native speakers tend to focus on a limited number of shibboleths as markers of linguistic or social identity. The fast and dirty nature of some early linguistic studies in Native North America has also promulgated certain misconceptions concerning dialect differentiation, and linguists are continuing their efforts to document basic regional dialect differences. Regular differences in vocabulary and pronunciation are still the starting points for this effort.

Sound and Lexicon

The dialects of Sioux (Dakota), which are spoken in the Great Plains of the United States and Canada, form a well-established dialect continuum. This continuum has been mistakenly broken into three distinct dialects based on a regular pronunciation difference that is apparent in the word for 'Indian': *lak^hota* – Teton, *dak^hota* – Santee, Sisseton, *nak^hota* – Yankton-Yanktonai, Assiniboine, and Stoney. Based on such linguistic division, natives have sometimes categorized each other along the same lines. A cursory examination, however, reveals this categorization to be inaccurate. Parks and DeMallie's report of their recent dialect survey revealed considerably more phonological variation for /l, d, n/ among the dialects, such as in the diminutive suffix and the word for 'little' (see Table 1). A better phonological diagnostic is in fact the pronunciation of the first consonant in a cluster when the second is a sonorant, as in *hnayā* 'cheat.' They concluded that phonologically Yankton and Yanktonai (Minnesota, Nebraska, Saskatchewan), rather than being n-dialects, are closer to Santee-Sisseton (Nebraska, Minnesota, North Dakota, Manitoba, Saskatchewan) than to Assiniboine (Saskatchewan, Montana). Stoney (Alberta), because of its vast difference phonologically and lexically and the influence of Cree (Algonquian) grammar, is a separate language from the others in the continuum.

Grammar

Although traditional phonological comparisons of lexicon and morphological affixes are indispensable for establishing dialect differences, Valentine has recently asserted that the incorporating nature of many indigenous languages offers a unique opportunity to trace variation within a dialect continuum and better understand similarities. The eight dialects of Ojibwe (Algonquian), spoken from Quebec to Saskatchewan, demonstrate considerable lexical variation because of the complex structure of their incorporating verbs. Verbs contain at least three basic slots: initial, medial, and final. For one verb, both the structural

components filling these slots and lexical variation within the slots may occur. Among the Ontario dialects, the animate intransitive body part verb *ozhaa-washkoshkiinzbigwe* 'have blue eyes' (Lake Nipigon) differs considerably although the initial, 'blue/green,' is the same (see Table 2). Tracing such complexity of variation is a fruitful direction for languages that become increasingly better documented.

Intergenerational Variation and New Language

Numerous linguists have commented on the difficulty of creating adequate descriptive grammars and dictionaries for languages, partly due to the amount of variation that exists between regional dialects and due to variable use, but also because of intergenerational shift. Such complexity is magnified in contact language situations, where intergenerational language attrition can be quite dramatic. Linguists of Boas's generation encouraged researchers to work with older, more expert speakers in order to have good comparative data, but as language attrition, maintenance, and revitalization have become part of the accepted reality of most linguists working in the Americas, focus has shifted to better understand variation caused by language attrition. It is now a type of variation not only associated with language death, but as vital to accounting for language revitalization.

In a comparatively early study, Hill traced the grammatical reductions in two southern California languages, Luiseño and Cupeño (Uto-Aztecan), demonstrating that in a 40-year period, speakers' grammar became increasingly more predictable as they displayed less subordination and came to favor shorter sentences. Cook, however, has observed the double nature of variation induced by language attrition: it tends to reduce the phonemic inventory, grammar, and stylistic options for individual speakers, but simultaneously the language on the whole exhibits more variation because the idiolects of semispeakers display a range of such reductions. Instead of speakers 'reducing' linguistic variation, they are in fact

Table 1 Sioux dialects

| | | | | | |
|-------------|---------------------------|--------------|------------------|------------|--------------|
| Teton | <i>lak^hota</i> | -la | <i>tjistila</i> | <i>gla</i> | <i>gnayā</i> |
| Santee | <i>dak^hota</i> | -da, -dā | <i>tjistina</i> | <i>hda</i> | <i>hnayā</i> |
| Sisseton | <i>dak^hota</i> | -na | <i>tjistina</i> | <i>hda</i> | <i>hnayā</i> |
| Yankton | <i>dak^hota</i> | -na | <i>tjistjina</i> | <i>kda</i> | <i>knayā</i> |
| Yanktonai | <i>dak^hota</i> | -na | <i>tjistjina</i> | <i>gda</i> | <i>knayā</i> |
| Assiniboine | <i>nak^hota</i> | -na | <i>tjusun</i> | <i>kna</i> | <i>knayā</i> |
| Stoney | <i>nak^hoda</i> | -n | <i>tjusin</i> | <i>hna</i> | <i>hnā</i> |
| | 'Indian' | 'diminutive' | 'little' | 'go home' | 'cheat' |

(Adapted from Parks and DeMallie, 1992.)

Table 2 Polysynthetic variation in Ojibwe

| Dialect/Location | Initial | Classificatory medial | | Body part medial | Final |
|------------------|------------|-----------------------|--------|------------------|----------|
| | 'blue' | augment | liquid | 'eye' | act.intr |
| Wikwemiking | o3a:wafkw- | -a: | -gam- | -i:ngw- | -e |
| Lake Nipigon | o3a:wafkw- | | | -fki:n3igw- | -e |
| White Dog | o3a:wafkw- | | -gam- | -fki:n3igw- | -e |
| Northern Ojibwe | o3a:wafkw- | -a: | -gam- | -d3a:b- | -i |
| Severn Ojibwe | o3a:wafkw- | | | -d3a:b- | -i |

(Adapted from Valentine, 2002: 91.)

caught at various stages of incomplete language acquisition. Cook maintained that the younger speakers of Sarcee (Sarsi) (Athabaskan) and Chipewyan (Algonquian) in western and northern Canada are speakers of 'dying' languages and therefore have a wide variety of innovations, archaisms, and phonological idiosyncrasies.

There is a fine ideological line between the speakers of a 'dying' language and those of a 'new' or 'revitalized' language although they may demonstrate the same phonological and structural reductions in addition to the variation typical of nonnative speakers. Recently, in a standard study of the regional dialects of Straits Salish (Salishan) spoken north and west of Puget Sound in Washington State and British Columbia, Montler grouped the dialects into two separate languages based on their nonmutual intelligibility: Klallam with three dialects and Northern Straits – a continuum of Sooke, Songish, Saanich, Lummi, Samish, and Semiahmoo. This classification is largely based on differences and similarities in lexical items, phonology, person marking and reduplication, demonstratives, and second-position enclitics. However, in Northern Straits, intergenerational variation within one family is as great as any variation from one dialect to the other. For instance, younger Saanich speakers tend toward periphrasis to express diminutives rather than reduplicating by adding */-C₁ə-/* after the first stressed syllable: */məmím'ən sɣénət/* 'small stone' instead of */sɣənɣéʔnət/*. When younger speakers do use reduplication, they display the greater individual variation described by Cook by either regularizing (as seen below) or exaggerating through additions of too many syllables:

Younger: *ʔiləqsən* 'point of land' *ʔəʔiʔləqsən* 'small point of land'
 Older: *ʔiləqsən* 'point of land' *ʔəʔələqsən* 'small point of land'

The greatest amount of innovation is seen, however, among new speakers of revitalized languages – New Lummi, New Saanich, and New Klallam (Clallam) – in which some speakers are very fluent. These varieties are growing through language revitalization efforts, and display even more of a tendency

to periphrastic structures as lexical suffixes are eliminated and verb paradigms are leveled. Phonologically, there is extreme change with the loss of glottalization on sonorants and some obstruents, vowel epenthesis in some clusters, and a shift toward more English-like sounds: */q/* becoming */k/* and */tɬ/* becoming */k'l/* or */kl/*. These are not symptoms of language death, but endemic of revitalization efforts as monolingual English speakers acquire an ancestral language.

Personal Indicators in Speech

It is somewhat common throughout the world's languages to modify the speech of direct address to demonstrate respect by modifying vocabulary, or using morphological plurals or passives to create a 'distancing' effect. Similar modifications are made in addressing strangers or in-laws in several indigenous languages such as those of northern California and Navajo (Athabaskan). Such accommodations in the presence of addressees serve to indirectly index their respected position in a speech event. Some indigenous languages of the Americas extend such indexicals to allude to specific or personal information of speech participants or when describing others. This is accomplished through a variety of manipulations of sound symbolism, lexicon, morphology, and phonology. The kinds of information expressed include specific physical characteristics or types of speech associated with certain characters in folktales, affection displayed toward younger or cute addressees, and the gender of the speaker and/or addressee.

Folk Characters and Abnormalities

Characters in folktales such as Coyote, Grizzly Bear, Mountain Lion, or Rabbit often have specific speech styles such as lisps, consonant substitutions, and frontings, or prefixing of certain sounds (see 'Special language' in Mithun, 1999 for a detailed overview). In Takelma (Oregon isolate), Dell Hymes argued such features are not used across the board but expressively to associate specific contextual qualities with the folk characters. Although Grizzly Bear speech in Takelma is well-known for */h/*-prefixing, this is only

done to indicate coarseness, stupidity, disdain, and distance. Likewise, indicators of physical or personality characteristics of people are not obligatory. Nootka (Wakashan/Vancouver Island), Quileut (Quileute) (Chimakuan/Washington), and purportedly a number of other northern Pacific coast languages oftentimes express negative personal characteristics such as greed, shortness, fatness, lameness, eye problems, or left-handedness through the addition of a morpheme and what Sapir calls “consonantal play.” For instance, an augmentative suffix */-aq/* is used for fat people: */haʔokwithak/* ‘did you eat?’ versus */haʔokwaqithak/* ‘did you eat, fatty?’ The diminutive suffix */-ʔis/* is added to forms referring to crossed or squint eyes, and all sibilants are converted to corresponding lateral forms: */qʷísmahl/* ‘he does so’ versus */qʷít-ʔít-mahl/* ‘he does so, weak eyes.’ Similar ‘sore eye’ speech without the diminutive is used to represent the folktale characters Deer and Mink.

Caretaker Language and Baby Talk

Similarly, caretaker language and baby talk is present in most languages of the northwest coast such as in the consonant reductions and the addition of diminutive */-ʔis/* in Nootka. Many, though not all, languages of North America have some modifications to indicate affectionate speech to young children. Such talk is characterized by varying degrees of modification of adult forms, such as morphological diminutives and reduplication, lexical substitutions, and phonological substitutions. The latter may involve a phonological shift such as palatalization of stop consonants, e.g. */t/* becomes */tʃ/* to indicate affection. Through a sort of diminutive sound symbolism as in Omaha-Ponca, grandmother language indicates affection, cuteness, or a minimized threat of a potential danger.

In other instances, there is a reduced phonological inventory because speakers perceive it to be easier for children. Cocopa, an Arizona Yuman language, has quite complex caretaker language with at least seven different phonological rules that adult women apply to varying extents when addressing children and adolescent girls. Consonants are reduced from the adult inventory of 23 to 12 largely through elimination of secondary articulations and place. For instance, */k, kʷ, q, qʷ/* become */k/*, and */l, lʷ, r/* become */l/*. Alternatively, adults may palatalize all dental, alveolar, and palatal consonants. Next, the consonant before the root syllable is replaced with a */v/*-sound although it is not a member of the adult Cocopa inventory. Consonants and sometimes vowels that are not part of the root get dropped; an */s/*-suffix is sometimes added to the end of the word; nonroot long vowels

are shortened, and a diminutive */n-/* prefix is added before the first consonant of the root. Men also produce the diminutive form. These changes and their variable applications create words in baby talk that can be quite complex for children to acquire in addition to their acquisition of adultlike language:

| Adult speech | Baby talk | |
|-----------------------------|--------------------------|--------------------------|
| <i>/kʷanʔuk/</i> | <i>/kanvúk/</i> | ‘baby’ |
| <i>/lu:p mu:kʷi: kmyul/</i> | <i>/lu:p unvi:s yul/</i> | ‘what will you buy’ |
| <i>/umits wa:ya:ts/</i> | <i>/unvít anya:tl/</i> | ‘she goes around crying’ |

Gender

Gender indexicals are typically phonological, morphological, or lexical variants used to indicate the gender of the speaker, the addressee, or both simultaneously. These continue to receive descriptive attention in Native North America, as linguists re-examine previously gathered data and find new examples. Despite early reports to the contrary, such as Haas’s article ‘Men’s and women’s speech in Koasati,’ the use of such indicators has rarely been found to be categorical. However, recent language attrition and the commensurate reduction of stylistic variation may in some circumstances have caused a greater regularity of use than was previously apparent.

Phonological Indicators Phonological gender differences have been proposed for a number of languages and exist in the length of forms, degree of nasalization, and in consonant substitution. There appears to be little crosscultural regularity in North America concerning which pronunciations will be more indicative of men than women. Haas first proposed that several Muskogean languages, specifically Koasati, indicated the male sex of the speaker, by pronunciation differences such as a final */-s/* in men’s forms: */lakawtakko/* (women) versus */lakawtakkos/* (men) ‘I’m not lifting it.’ Using Haas’s original field notes, historical records, comparative data, and his current fieldwork, Kimball argued that the phonological difference Haas analyzed for Koasati and proposed for related languages was actually a choice in whether to use */-s/* (currently pronounced [ʃ]) ‘sentence-final narrative particle’ to supplant phrase-terminal nasalization. He also noted that this feature was due less to a speaker’s gender than to his or her authority in the society as both sexes used the form. Yana (northern California) men’s public speaking to both men and women was originally described by Sapir as male-to-male speech. Because of the additional syllable, ‘men’s forms’ are lengthier, more archaic, and evocative of speech associated with elevated ritual language and indirectly with gender.

| | | |
|----------------------|-----------------------------------|---------------|
| Man to man | Other | |
| /ʔau-ʔmidʒa | /ʔau-ʔitʃ ^h / | 'my fire' |
| /ʔi-na/ | /ʔi/ | 'tree, stick' |
| /p ^h adi/ | /p ^h at ^h / | 'place' |

Likewise, in ritual Laguna (Keres) *kiva* speech associated with older men, regular vowel lengthening occurs in 10 'cue' words indicating the emotional stance of the speaker toward a proposition.

| | | |
|---------|----------|--------------|
| Female | Male | |
| /amú'u/ | /amúu'u/ | 'love, pity' |
| /ayá'a/ | /ayáa'a/ | 'discomfort' |
| /imí'i/ | /imíi'i/ | 'fear/shy' |

Several languages display slightly more nasalized forms for speech attributed by native speakers to women. In the early part of the 20th century, Inuktitut (Eskimo-Aleut) women on Baffin Island were observed to substitute nasals /m, n, ŋ, N/ for final voiceless stops /p, t, k, q/. Yet men also had variable nasal and non-nasal pronunciations. Among the Lakhota (Lakota) (Siouan) affective/illocutionary force indicators, two to three forms, such as /-yemā/ 'women's surprise,' are nasalized versions of men's – /-yewā/ 'men's surprise'. In contrast, emphatic questions in Yana required the enclitic /-gà/ for women and /-nà/ for men.

In Atsina (Gros Ventre) (Algonquian/Montana) both men and women tend to use 'women's' pronunciation' with nonnative speakers and children, but men typically front /k/ or /ky/ to /tʃ/, and women substitute /k/ or /ky/ for the more archaic /ty/. Men have apparently borrowed the fronted sound /tʃ/ from their Arapaho (Algonquian) neighbors, and younger men are currently in the process of changing /ty/ to /tʃ/ as well.

Vocabulary and Morphology Interjections, common sayings, speech act indicators, and some kinship terms vary according to the gender of the speaker or addressee in a number of languages. It is common for men and women to have different interjections for expressions of surprise, fear, or bravery in North American languages. Less common is for markers of illocutionary force to be gendered. Several Siouan languages, however, possess verbal suffixes like those listed in Mandan (Siouan) below. Mandan is the only Siouan language in which such forms are required for every sentence, and which indexes only the gender of the addressee.

| Imperative Statement | | Interrogative | | |
|----------------------|--------|---------------|--|------------------|
| /-rã/ | /-ʔre/ | /-ʔrã/ | | female addressee |
| /-ta/ | /-ʔ/ | /-ʔa/ | | male addressee |

Other common expressions of politeness and greeting can become very salient indicators of gendered difference for native speakers. Although the forms in Table 3 do not occur in every sentence of the pueblo

Table 3 Pueblo Southwest gendered vocabulary

| Language (dialect) | Female | Male | Meaning |
|-----------------------------|-------------------------|---------------------------------------|------------------|
| Hopi (3 rd Mesa) | ʔask ^w ali | k ^w ak ^w ha(-y) | thank you |
| Tewa (Arizona) | kuna | kunda | thank you |
| Tiwa | hɛrkem | hawə | thank you |
| Acoma | náidrá | huw'ehé | thank you |
| Hopi | sónwayo | lóloma | it's beautiful |
| Acoma | an ^y umé:c'a | ʔanyi:c'e | it's beautiful |
| Tewa (Arizona) | ʔasagi | sag ^y wo' | it's beautiful |
| Hopi | | taʔa | yes |
| Tewa (Arizona) | hā: | hoy | yes |
| Tewa (Rio Grande) | hoy | hāman | yes |
| Acoma | hée | haí | answer to a call |
| Hopi | yá:sayoqu | hósqaya | be huge |
| Tewa (Arizona) | -ʔáyyá | -ʔóyyó | be good |

(Adapted from Maring, 1975; Kroskrity, 1983: 89; Sims and Valiquette, 1989.)

southwest languages, there is a strong ideology of gender-differentiated speech.

The Meanings of Gendered Forms Lists of gender differences out of context give a very minimal sense of the meaningful causes of such variation. The question of how such forms get connected to gender as they index different genres, meanings, and speech acts will lead us further to understanding the meaning of variation in specific languages (see Trechter, 1999 for such an analysis in Lakhota). Indeed, any number of linguistic forms could theoretically be linked to gender, but very few are. For example, Hill and Zepeda have recently extended the types of analysis indicative of gender differentiation to the use of the ingressive pulmonic airstream by Tohono O'odham (Uto-Aztecan/Arizona) women. The women's sucking in of breath coordinates with discourse features that index closeness and involvement in conversation, which already hold a meaningful culture value for these women. Only through contextual analysis coordinated with speakers' ideologies regarding gendered meaning are such linguistic markers better understood.

Stylistic Variation

In a 1927 article, 'Literate and illiterate speech,' Bloomfield assessed Menominee (Menomini) (Algonquian) speakers in Wisconsin, rating their grammatical and lexical capabilities. He described three different ways of saying 'What are you laughing at?' – /wékiʔ wéh-ayéniyan/, /wékiʔ aya:yó:sinamal/, /tá:niʔ wehtá:hpiyan/ – on a continuum from 'illiterate, childish, stupid' to 'normal' to 'elevated, poetic and archaizing.' The last type of language seemed to be associated with doctors and shamans, and was

characterized by long vowels in unusual contexts, archaisms, and metaphorical vocabulary. 'Bad' speakers, on the other hand, anglicized their pronunciation and did not keep long and short vowels distinct. They also forgot to use the obviative (a form that distinguishes between two different third persons) and the quotative in narratives. Although Bloomfield placed most speakers under the age of 40 in the illiterate group, this was not absolute. He surmised that English interfered with a speaker's ability to control a variety of styles, but he also placed some monolingual speakers of Menominee in the illiterate group. Regrettably, Bloomfield stated that the differences in such styles of speech permeated every aspect of the lexicon and grammar to the extent that description would be impossible. It is difficult to know if Bloomfield was capturing the reduction of stylistic variation or the prejudice of a few speakers.

The Content of Style

Elevated, ceremonial, or shamanistic speech styles are recognized among many languages (Wintu, Lakota, languages of the pueblos in the southwest). On the other hand, styles of oratory in many California and west coast languages (Nootka, Tübatulabal, Patwin, Pomo, Yokuts) is reported as having been 'forced,' 'jerky,' with 'short sentences.' For an excellent comparison of such speech styles see Miller and Silver (1997). In cultures where mastery of ceremonial speech is highly emphasized, speakers must have specific abilities and be specially trained to acquire the forms. Such registers thus become less indicative of the personal or individual identity of a speaker than the genre or speech event which she or he performs.

Chafe, for instance, described three different styles of verbal performance in Seneca (Iroquoian/New York): normal conversation, preaching, and chanting. These genres differ in a number of ways: prosodically, and in the degree of formulaicity, sentence grammar, and epistemic stance. Normal conversation has greater freedom, relies less on memorized phrases, circumlocutions, or archaic language, is more fragmented because of multiple participants and lack of planning, and requires speakers as direct conveyors of original content to state the extent to which they are sure through the use of varied evidentials. Chafe demonstrated that as speakers move to preaching and chanting, their intonation contours become increasingly monotone although the sentences are lengthier and more packed with information. Preaching exhibits few evidential markers. Indeed, as the speaker becomes less individually and creatively responsible for the content of utterances, chanting contains a number of markers of certainty.

Although the definition of ritual speech requires a certain compartmentalization of the genre, there may be ideological extensions of ritual genre into everyday life to the extent that it may also affect speakers' sense of what more elevated language sounds like. From Bloomfield's description, some such linguistic ideology may have affected his definition of elevated speech in Menominee with its elongated vowels and metaphorical vocabulary. To some extent, formulaic speeches, such as the public grievance chants in Hopi (Uto-Aztecan), may have some general similarity in form to more esoteric styles. Kroskrity argued that the speech of public announcements in Arizona Tewa (Kiowa-Tanoan), although necessarily separate from the highly ritualistic *kiva* speech, follows many of the same strictures through its de-emphasis on the personality of the announcer, the formulaic structure of the announcing chant, and the unusual but predictable rising intonation throughout the three verses. To be 'valid' forms, such chants must be notifications of community events and announced by men from rooftops. More recently women have both followed the form and innovated by announcing events such as yard sales themselves.

Folktales as Verse

Since the 1970s, a number of linguistic anthropologists such as Tedlock and the Hymeses have argued that the structure, form, language, and meaning of oral, indigenous folktales are better understood if they are regarded as poetry and performance rather than prose. Tedlock, for instance, has sought to capture visually variations in loudness and prosodic features such as intonation and rhetorical lengthening through use of larger type, movement of the type up and down on the page, and the repetition of letters for lengthening. Dell Hymes saw the 'pattern numbers,' those numbers that are favored mythically by certain cultures for organizing, as important for understanding if not for reversing previously collected folktales. One of the most famous of such analyses using this approach is Virginia Hymes's work on the 'Raven Myth' in Warm Springs Sahaptin (Tenino) (Sahaptian/Oregon). Paying special attention to the intersection of prosodic cues such as changes in voice quality or pitch and rhetorical vowel lengthening indicative of first lines, with parallelism in the use of different grammatical particles and time words, Hymes was able to divide the narration into verses. Interestingly, a pattern of five recurred throughout. Events were repeated five times, and there were often five lines in a verse and five verses in a stanza. The pattern of three was also important. Despite the variation in some folktales and creative interpretation needed for such

work, Woodbury saw such work as getting to the grammar of discourse. By understanding the similarity of forms within one oral genre, one thereby begins to see its variation from others.

Variation in Native North America has never been limited to regional dialects or the expression of cultural, ethnic, or gender identities. Linguists encountering unexpected and unfamiliar linguistic structures and expressions in North America have allowed a broad definition of variation in order to capture both the sense of complexity in the languages they described and their expressive capabilities. As today's linguists and Native peoples continue to work in a context of language attrition and potential revitalization, maintaining a broad understanding of variation in both the older forms and encountering variation in new contexts will be the challenge.

Bibliography

- Baumann R & Sherzer J (eds.) (1974). *Explorations in the ethnography of speaking*. Cambridge: Cambridge University Press.
- Bloomfield L (1927). 'Literate and illiterate speech.' *American Speech* 2, 432–439.
- Boas F (1917). 'Introductory.' *International Journal of American Linguistics* 1, 1–8.
- Chafe W (1993). 'Seneca speaking styles and the location of authority.' In Hill J & Irvine J (eds.) *Responsibility and evidence in oral discourse*. Cambridge: Cambridge University Press. 72–87.
- Cook E (1989). 'Is phonology going haywire in dying languages? Phonological variations in Chipewyan and Sarcee.' *Language in Society* 18, 235–255.
- Crawford J (1978). 'More on Cocopa baby talk.' *International Journal of American Linguistics* 44, 17–23.
- Greenfield P (1999). 'Some special morphological characteristics of the White Mountain dialect of Apachean.' *Anthropological Linguistics* 41, 356–381.
- Haas M (1944). 'Men's and women's speech in Koasati.' *Language* 20, 142–149.
- Hill J (1978). 'Language death, language contact, and language evolution.' In McCormack W & Wurm S (eds.) *Approaches to language*. The Hague: Mouton. 45–78.
- Hill J & Zepeda O (1999). 'Language, gender, and biology: pulmonic ingressive airstream in women's speech in Tohono O'odham.' *Southwest Journal of Linguistics* 18, 15–40.
- Hymes D (ed.) (1981). *In vain I tried to tell you: essays in Native American ethnopoetics*. Philadelphia: University of Pennsylvania Press.
- Hymes V (1987). 'Warm Springs Sahaptin narrative analysis.' In Sherzer J & Woodbury A (eds.) *Native American discourse: poetics and rhetoric*. Cambridge: Cambridge University Press.
- Kess J & Kess A (1986). 'On Nootka baby talk.' *International Journal of American Linguistics* 52, 201–211.
- Kimball G (1987). 'Men's and women's speech in Koasati: a reappraisal.' *International Journal of American Linguistics* 53, 30–38.
- Kinkade D & Mattina A (1996). 'Discourse.' In Sturtevant W & Goddard I (eds.). 245–274.
- Kroskrity P (1983). 'On male and female speech in the Pueblo Southwest.' *International Journal of American Linguistics* 49, 88–91.
- Kroskrity P (1992). 'Arizona Tewa Kiva speech as a manifestation of linguistics ideology.' *Pragmatics* 2, 297–309.
- Maring J (1975). 'Speech variation in Acoma Keresan.' In Kinkade D, Hale K L & Werner O (eds.) *Linguistics and anthropology in honor of C. F. Voegelin*. Lisse, Belgium: Peter de Ridder. 473–485.
- Miller W (1996). 'The ethnography of speaking.' In Sturtevant W & Goddard I (eds.). 222–244.
- Miller W & Silver S (1997). *American Indian languages: cultural and social contexts*. Tucson: University of Arizona Press.
- Mithun M (1999). *The languages of Native North America*. Cambridge: Cambridge University Press.
- Montler T (1999). 'Language and dialect variation in Straits Salishan.' *Anthropological Linguistics* 41, 463–501.
- Parks D & DeMallie R (1992). 'Sioux, Assiniboine, and Stoney dialects: a classification.' *Anthropological Linguistics* 34, 233–249.
- Sapir E (1949). 'Abnormal types of speech in Nootka.' In Mandelbaum D (ed.) *The collected writings of Edward Sapir*. Berkeley: University of California Press. 179–196.
- Sims C & Valiquette H (1989). 'More on male and female speech in (Acoma and Laguna) Keresan.' *International Journal of American Linguistics* 55, 162–166.
- Sturtevant W & Goddard I (eds.) (1996). *Handbook of North American Indians 17: Language*. Washington, DC: Smithsonian Institution Press.
- Tedlock D (1983). 'On the translation of style in oral narrative.' In Swann B (ed.) *Essays on Native American oral literature*. Berkeley: University of California Press. 57–77.
- Trechter S (1999). 'Contextualizing the exotic few: gender and language in Lakhota.' In Bucholtz M, Liang A & Sutton L (eds.) *Reinventing identities*. New York: Oxford University Press. 101–122.
- Valentine R (2002). 'Variation in body-part verbs in Ojibwe dialects.' *International Journal of American Linguistics* 68, 81–119.

Navajo

H Landar

© 2006 Elsevier Ltd. All rights reserved.

Navajo is an Athapaskan (Apachean) language, with perhaps 200 000 speakers in Arizona, New Mexico, and Utah. Vowels /i e o a/ are single, clustered, nasalized (ǣ), and of low or high pitch. Syllables have strong, medium, or ordinary stress. Junctures are rising, falling, and sustained. Voiceless stops are labial, apical, and velar /p t k/ written *b, d, g*; labial and palatal sonorants are /w y/; voiceless nonlabial stops /t k/ and affricate clusters with /t/ have aspirated counterparts, e.g., /th kh/ written *t, k*, /tsh tšh/ written *c, č*. Aspirated segments are allophones of /h/; checked segments, of /ʔ/. Thus /tšʔ/ is the componential reduction of č. Voiceless nonfaucal fricatives have voiced homorganic counterparts, apical /s z/, palatal /š ž/, velar /x ɣ/, lateral /l ɭ/. Nasals are /m n/. Navajo lacks a Proto-Athapaskan series of labialized consonants and contrast of front and back velar consonants.

Navajo is a SOV language. Word classes are nouns, postpositions (adjuncts), verbs, and particles. Their functions define syntactic classes, nominals, adjunctives, verbals, and relationals. Pronominal prefixes of nouns and postpositions are similar in form but not in meaning; those of nouns are possessive, whereas those of postpositions translate as datives and accusatives, with adverbial stems. Compare *ší* 'I' (independent subject pronoun), *ši-má* 'my mother' (noun), *ši-tšʔ* 'me-towards' (postposition); *ní* 'you (SG)', *ní-lʔ* 'your horse(s)', *nitšʔ* 'toward you'; *bí* 'he (she, it, they)', *b-ádi* 'his older sister(s)', *b-aa* 'to him.'

Verb prefixes take 10 positions before a stem: adverbial, iterative, plural, objective, deictic, aspectual, modal, perfective, subjective, and classifier. Stem shapes vary for modes (imperfective, perfective, optative, and others) and aspects (momentaneous, continuative,

and others); some stems, however, are invariable. Neuter verbs are intransitive; active verbs include agent and patient; third persons and other prefixes may have a zero (∅) form. Thus *ní-ʔaah* 'you (SG) move it (as a compact or round object)'; cf. *š-aa ní-ʔaah* 'give it to me'; that is, 'to me you move it as a compact object.' The stem *-ʔaah* belongs to a classificatory gender system for movement or handling of various objects (e.g., a container with contents, a living being, a flat flexible object, a thin rigid object, a ropelike object, and others). Navajo shares verbal patterns with other Athapaskan languages but has lost a rich Proto-Athapaskan system of prefixation and suffixation for aspect and mode.

Sentences are transformed principally by rules for moving words or inserting enclitics (a subclass of particles). The enclitic *íš* 'is it?' makes one type of question: *ʔaškii* ('boy') *ʔatʔééd* ('girl') *yiztsqs* ('he kissed her') *íš* ('is it?'), 'did the boy kiss the girl?' Negatives are generated commonly by *doo ... da*: *doo yiztsqs da* 'he did not kiss her.' The prefix *yi-* alternates with *bi-* to indicate the third person object when the subject is a third person (∅ in position (i)). *Yi-* and *bi-* involve a hierarchy of animacy for some speakers; they may preserve the Proto-Athapaskan–Eyak semitransitive category identified by Krauss (1965). Often *yi-* shows that an action is controlled or semitransitive while *bi-* shows full transitivity on a foregrounded object. Thus object advancement from *ʔaškii ʔatʔééd yiztsqs* 'the boy kissed the girl' to *ʔatʔééd ʔaškii biztsqs* 'the girl was kissed by the boy' foregrounds NP₂ and cancels semitransitivity.

Bibliography

Krauss M E (1965). 'Eyak: a preliminary report.' *Canadian Journal of Linguistics* 10, 167–187.

Nenets

F Ackerman, University of California, San Diego, CA, USA

T Salminen, University of Helsinki, Helsinki, Finland

© 2006 Elsevier Ltd. All rights reserved.

Nenets is a subbranch of the Samoyed branch of the Uralic family comprising two closely related but distinct languages, Forest Nenets (FN) and Tundra

Nenets (TN). Tundra Nenets is spoken by nearly 30 000 people across the vast tundra zone of Arctic Russia and northwestern Siberia, while Forest Nenets has perhaps 1500 speakers along the Pur, Agan, Lyamin, and Nadym river basins in northwestern Siberia. A clear majority of the speakers are proficient in Russian, and in the European part of the Tundra Nenets territory in particular, the native language is in these days rarely transmitted to younger generations.

In addition to Russian, Tundra Nenets has had contacts especially with Komi and Northern Khanty, and Forest Nenets has been greatly influenced by Eastern Khanty.

Besides Nenets, the Samoyed branch includes Nganasan, Enets (Forest Enets and Tundra Enets), Yurats, Selkup (Northern Selkup, Central Selkup, and Southern Selkup), Kamas, and Mator; of these, Yurats, Kamas, and Mator are extinct, the Enets languages as well as Central Selkup and Southern Selkup are critically endangered; Nganasan is still spoken by approximately 500 people and Northern Selkup by 1500. Samoyed is the easternmost branch of the Uralic family; the other branches are Khanty, Mansi, Hungarian, Permian, Mari, Mordvin, Finnic, and Saami.

The Nenets languages are synthetic, agglutinating with some fusion and, in Forest Nenets, metaphony, morphophonologically complex, suffixing and predominantly verb-final.

The vowel system of Tundra Nenets in the first syllable includes nine vowels differing in both quality and quantity (one short vowel marked with \emptyset in phonological transcription, five basic vowels, $i\ e\ a\ o\ u$, a mixed [diphthongoid] vowel æ , and two long vowels, $i\acute{u}$; in unstressed syllables, a schwa, ° , typically realized as extra lengthening of the preceding segment, occurs in addition to the five basic vowels. The Forest Nenets vowel system has been restructured after the Eastern Khanty model and consists of stressed syllables of six long vowels, $i\ e\ \text{ä}\ a\ o\ u$, and four short vowels, $i\ \text{ä}\ \hat{a}\ \hat{u}$ (corresponding to $i\ \text{ä}\ a\ u$); in unstressed syllables, only a schwa ° and $i\ a\ u$ are possible. The stress is not contrastive but falls on nonfinal odd or pre- and postschwa syllables. A feature affecting both consonants and vowels is palatalization: the traditional formulation is that vowels have back vs. front allophones after nonpalatalized vs. palatalized consonants, but palatality (marked with y between a consonant and vowel in phonological transcription) can also be understood as a suprasegmental feature with a CV sequence under its scope. The consonant system of Tundra Nenets consists of 26 units (up to 31 in dialects); in Forest Nenets there are 24 consonants. Both systems include a velar nasal (ng) and a velar fricative (x); in Forest Nenets, vibrants have changed to fricolaterals (lh) under the Eastern Khanty influence; in Tundra Nenets, there are affricates (c) that have developed from consonant clusters still retained in Forest Nenets; both languages have a glottal stop marked with q or, in Tundra Nenets, h in case it has nasal sandhi alternants. The above figures include palatalized consonants, which in Tundra Nenets are only contrastive in the labial and dental series, while in Forest Nenets, there are palatalized velars as well.

An old phonotactic peculiarity of Nenets is the lack of initial vowels: this is now relaxed in most varieties, but in the Central dialects of Tundra Nenets the principle is still fully alive and is even reflected in recent Russian loanwords such as *ngarmiya* ‘army.’ In Tundra Nenets, there is a sandhi system affecting both the final consonant of the preceding word and the initial consonant of the following one, for instance, *nyeh xøn°* ‘woman’s sledge’ is transformed to *nyeng_køn°*, *pyiq xøn°* ‘sledge for wood’ to *pyi_køn°*, and *ngarka* to ‘big lake’ to *ngarka_do* by sandhi.

Nouns distinguish seven cases: nominative, accusative, and genitive are the grammatical cases that in their basic functions denote subject, object, and possessor; dative, locative, ablative, and prosecutive (‘through, along, by’) constitute the local cases. There are three numbers, singular, dual, and plural, but there is a gap in the nominal paradigm in that the local cases do not combine with the dual number, the respective meanings being expressed by postpositional phrases. The inflection of personal pronouns follows a distinct pattern, and their local cases are also replaced with forms of postpositions. Besides absolute declension, the nominal inflection includes possessive as well as predestinative (‘for’) forms, e.g., FN *wyiq* ‘water’: *wyiqj°* ‘my water’: *wyiqtâj°* ‘water for me.’ The postpositions are typically inflected in local cases and have possessive forms as well, e.g., FN ablative *ngîlh°tâj°* ‘from under me’ or prosecutive *pumnantung* ‘along their tracks.’ In predicative position, nouns agree with the subject employing the same personal suffixes (but not showing the other inflectional peculiarities) as intransitive verbs, e.g., TN *lúca* ‘Russian’: *lúcad°m* ‘I am a Russian.’

Verbs have numerous grammatical categories, covering person, number, tense, and mood. The number of moods is large, in Tundra Nenets up to 18, making it possible to express various levels of probability and necessity morphologically; the imperative and optative moods employ sets of personal suffixes different from other moods. Perfective vs. imperfective aspect is an inherent feature of a verb, and aspectual pairs are created through derivational morphology. The tense is expressed by two distinct systems: first, there is an opposition between unmarked basic tense and suffixally marked future and habitive tenses; second, there is unmarked aorist vs. preterite marked by a suffix that morphotactically follows the personal suffix; it is possible to combine the two tense systems, e.g., TN *xada-* ‘kill’: aorist *xadaøw°* ‘I killed it (just now)’: preterite *xada°wøsy°* ‘I killed it (earlier)’: future aorist *xadangkuw°* ‘I am going to kill it’: future preterite *xadangkuwøsy°* ‘I was going to kill it.’ As seen from the examples, the basic aorist refers to

immediate past in case of perfective verbs such as 'kill,' whereas the aorist of imperfective verbs simply expresses present, e.g., *nyoda-* 'follow': *nyodaow°* 'I am following it.' A specific grammatical category in Nenets is known as conjugation: it covers the opposition between subjective forms used when the object is focused and objective forms referring to previously known or omitted objects, e.g., TN *tim xadaod°m* 'I killed a/the reindeer (and not another animal)' vs. *tim xadaow°* 'I killed the reindeer (instead of doing something else to it)'; in the objective conjugation, the number of the object is expressed morphologically, e.g., *xadangax°yun°* 'I killed them (two)' vs. *xadeyøn°* 'I killed them (several)'; furthermore, there are reflexive forms that either contrast with forms with a transitive meaning, e.g., *tonta-* 'cover': objective *tonta°da* '(s)he covered it': reflexive *tontey°q* 'it got covered,' or constitute the only finite forms of a lexical verb, typically expressing sudden movement or change in state. The personal suffixes cannot generally be attached directly to the verbal stem, but they trigger a complex system of morphological subSTEMS.

There is a wide range of nonfinite verbal forms in Nenets, with an important function in embedded clauses (either independently or within postpositional phrases, often with switch-reference, whereby a nonfinite verb is marked differently depending on whether its subject is the same as, or differs from, that of the finite verb), as there are no conjunctions or relative pronouns. Negation is expressed by a negative auxiliary verb incorporating all categories of verbal inflection followed by a specific connegative form of a lexical verb, e.g., TN *nyix°yun° xadaq* 'I did not kill them (two)'; since the nominal paradigm lacks a connegative, negative nominal predicates must incorporate a copula, e.g., TN *lucad°m nyid°m ngaq* 'I am not a Russian.'

Within the basic SOV word order of a transitive sentence, the adverbial phrases are typically placed as Time S Place/Recipient O Manner V, but any focused element can occur preverbally, and even postverbal constituents are possible in case of two morphologically or functionally similar phrases, e.g., FN *ngopk°na myatuqngaj° mâj° myaqk°naj°* 'we (two) live together in our tent,' where both *ngopk°na* 'together' and *mâj° myaqk°naj°* 'in our tent' are in the locative case. In imperative sentences, typically without an overt subject, the nominal object is in the nominative instead of the accusative. The personal pronouns, by contrast, employ their accusative forms even in imperative sentences, while in possessive phrases with a morphologically marked possessed noun they appear, if not omitted, in the nominative

rather than in the genitive. Agreement within a nominal phrase is possible in number when the nonsingularity of the noun is more definite, and in relative clauses possessive agreement also occurs, e.g., TN *metyida wadyida* 'note the words he uses', cf. *meta* imperfective participle of 'use,' *wada* 'word.'

Both Nenets languages are endangered, but there are major differences between localities in language use. Tundra Nenets has a literary language deriving from the 1930s used in semiregular book printing and having a limited presence in schools and the press, while Forest Nenets remained unwritten until the 1990s, when a primer and a school dictionary appeared. In the areas where the languages remain vigorous, oral literature, including tales, stories, and riddles as well as epic, lyric, and personal songs, is also flourishing (Castrén and Lehtisalo, 1940; Lehtisalo, 1947; Kupriyanova, 1965; Tereshchenko, 1990; Niemi, 1998). The traditional way of life based on reindeer husbandry or fishing (Khomich, 1995) continues to be appreciated by many Nenets as long as oil and gas excavations do not entirely destroy their lands and the authorities do not force them to relocate (Golovnev and Osherenko, 1999).

For a small indigenous language, Tundra Nenets is reasonably well studied, especially with regard to its phonology and morphology (Castrén, 1854; Tereshchenko 1947, 1956; Décsy, 1966; Janhunen, 1986; Salminen, 1997, 1998a) and lexicon (Pyrerka and Tereshchenko, 1948; Lehtisalo, 1956 [covering both Nenets languages]; Tereshchenko, 1965), while there is only one monograph devoted to the syntax of the Samoyed languages in general (Tereshchenko, 1973). This article is mainly based on Salminen (1998b) as well as more recent field studies funded by the Endangered Languages Documentation Programme. Forest Nenets has been studied much less extensively than Tundra Nenets, with a couple of basic grammatical treatments published (Verbov, 1973; Sammallahiti, 1974).

Bibliography

- Castrén M A (1854 [1966]). *Grammatik der samojedischen Sprachen*. St Petersburg: Herausgegeben von Anton Schiefner. [Indiana University Publications, Uralic and Altaic Series 53; Bloomington & The Hague: Mouton.]
- Castrén M A & Lehtisalo T (1940). *Samojedische Volksdichtung*. Mémoires de la Société Finno-Ougrienne 83; Helsinki.
- Décsy G (1966). *Yurak Chrestomathy*. Indiana University Publications, Uralic and Altaic Series 50. Bloomington and The Hague: Mouton.

- Golovnev A V & Osherenko G (1999). *Siberian survival: the Nenets and their story*. Ithaca, NY: Cornell University Press.
- Janhunen J (1986). *Glottal stop in Nenets*. Mémoires de la Société Finno-Ougrienne 196; Helsinki.
- Khomich L V (1995). *Nency: ocherki tradicionnoy kul'tury*. Sankt-Peterburg: Russkiy dvor.
- Kupriyana Z N (1965). *È 9picheskie pesni nencev*. Moskva: Nauka.
- Lehtisalo T (1947). *Juraksamojedische Volksdichtung*. Mémoires de la Société Finno-Ougrienne 90; Helsinki.
- Lehtisalo T (1956). *Juraksamojedisches Wörterbuch*. Lexica Societatis Fenno-Ugricae 13; Helsinki.
- Niemi J (1998). *The Nenets songs: a structural analysis of text and melody*. Acta Universitatis Tamperensis 591; Tampere.
- Pyrerka A P & Tereshchenko N M (1948). *Russko-neneckiy slovar'*. Moskva: Ogiz-Gis.
- Salminen T (1997). *Tundra Nenets inflection*. Mémoires de la Société Finno-Ougrienne 227; Helsinki.
- Salminen T (1998a). *A morphological dictionary of Tundra Nenets*. Lexica Societatis Fenno-Ugricae 26; Helsinki.
- Salminen T (1998b). 'Nenets.' In Abondolo D (ed.) *The Uralic languages*. London: Routledge. 516–547.
- Sammallahti P (1974). *Material from Forest Nenets*. Castrenianum toimitteita 2; Helsinki.
- Tereshchenko N M (1947). *Ocherk grammatiki neneckogo (yurako-samoedskogo) yazyka*. Leningrad: Uchpedgiz.
- Tereshchenko N M (1956). *Materialy i issledovaniya po yazyku nencev*. Moskva and Leningrad: Izdatel'stvo Akademii nauk SSSR.
- Tereshchenko N M (1965). *Nenecko-russkiy slovar'*. Moskva: Sovetskaya ènciklopediya.
- Tereshchenko N M (1973). *Sintaksis samodiyskikh yazykov: prostoe predlozhenie*. Leningrad: Nauka.
- Tereshchenko N M (1990). *Neneckiy èpos*. Leningrad: Nauka.
- Verbov G D (1973). 'Dialekt lesnykh nencev.' In *ng1033 Samodiyskiy sbornik*. Novosibirsk. 3–190.

Nepali

M Hutt, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Nepali, a member of the Indo-Aryan group of languages, is the national language (*rāṣṭra bhāṣā*) of Nepal, the state language of Sikkim, and the sole language of most ethnic Nepali communities in Bhutan and northeast India. It was previously known as Khas Kurā (the speech of the Khas) or Gorkhālī. Nepali probably has about 17 000 000 mother tongue speakers, and is a vital second language for approximately 7 000 000 speakers of other Nepalese languages, many of which are Tibeto-Burman.

Nepali was introduced into the central Himalaya by immigrants who entered from the northwest before the 10th century. Its ascendancy over the other languages of the region is linked to a process of political domination and cultural assimilation. Written in the Devanāgarī script, its earliest records are 13th-century royal inscriptions from far western Nepal, though Nepali was rarely used for literary purposes until the 18th century, and its first major work, the *Nepālī Rāmāyaṇa* of Bhanubhakta Acharya, was written in the mid-19th century.

Among the other major Indo-Aryan languages, Hindi is Nepali's closest cousin, and many literate Nepali speakers are proficient in Hindi. However, in its everyday vocabulary Nepali preserves many Sanskrit and Sanskrit-derived words (e.g., *ghām* 'sun', *khukurā* 'chicken', *ritto* 'empty') that have

been displaced by Perso-Arabic loans in Hindi, and the Arabic and Perso-Arabic element of its lexicon is largely confined to law, war and weaponry, and governance and monarchy. Similarly, English loans are generally less common in Nepali than in Hindi, partly because Nepal was never colonized by the British.

Unlike Hindi, Nepali distinguishes between existential (*chanu*) and definitive (*hunu*) functions of the verb 'to be', e.g., *pānī ho?* 'is this water?', *pānī cha?* 'is there [any] water?' Like Bengali, it uses numeral classifiers, e.g., *tīnjanā mānche* 'three [-person] men', *tīnvaṭā mec* 'three [-object] chairs' and accords feminine gender only to female human nouns. It forms most plural nouns through the affixation of *harū*; and generally forms negative verbs by the adaptation of verb endings, e.g., *ma jānchu* 'I go', *ma jādina* 'I do not go'. Four honorific grades plus a royal honorific are available for personal pronouns. Clauses are commonly linked by the use of infinitives and participles and seldom by conjunctions: thus, *H. vah ādmī jo kal āyā* 'the man who came yesterday' is *Ne. hījo āeko mānche* 'the yesterday having-come man'.

Nepali lacks the H. phonemic distinction of /s/ /v/ from /s/ /b/. Regional dialects have been identified but dialectal variation is not strong. Nepali has influenced the Tibeto-Burman languages of the region more than it has been influenced by them (Sprigg, 1987), though there are a few loanwords from Newari, e.g., *jhyāl* 'window' and some features of syntax and intonation may reflect Tibeto-Burman influence.

Bibliography

- Clark T W (1969). 'Nepali and Pahāḍḍ.' In Sebeok T A (ed.) *Current Trends in Linguistics, vol. 5: Linguistics in South Asia*. Paris and the Hague: Mouton. 249–276.
- Hutt M J (1997). *Modern literary Nepali: an introductory reader*. Delhi: Oxford University Press.

- Sprigg R K (1987). 'The Nepali language with reference to its relationship with the Rai, Limbu and Lepcha Languages.' *Indian Linguistics* 48, 87–100.
- Turner R L (1931). *A comparative and etymological dictionary of the Nepali language*. London: Routledge and Kegan Paul.

Ngan'gi

N Reid, University of New England, Armidale, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Ngan'gi (Ngan'gikurunggurr) is an Australian Aboriginal language spoken in the Daly area several hundred kilometers to the southwest of Darwin. The language now has two significant dialectal variations: Ngan'gikurunggurr has about 150 speakers, and Ngen'giwumirri has about 50. These two dialects share about 89% of vocabulary, have nearly identical systems of complex verb morphology, and are mutually intelligible. For our purposes, we can treat them as a single language and give it the label 'Ngan'gi,' although speakers are careful to distinguish between them as characterizing two separate groups with distinct social identities. Ngan'gi is spoken predominantly in the townships of Nauiyu (formerly Daly River) and Peppimenarti, and the outstations that those two towns supply. Most people who speak Ngan'gi fluently are aged more than 50 years, and the children of Ngan'gi speakers now mostly learn Kriol as their first language. With this number and age profile of speakers, Ngan'gi is classified as a 'threatened' language, under real danger of extinction within a few decades.

Classification

The first classification of Daly region languages (Tryon, 1974) paired Ngan'gikurunggurr and Ngen'giwumirri as constituting the Tyemeri branch of the 'Daly Family,' with neither related to Murrinh-Patha, the neighboring language to the west. This understanding, of either no relationship or at best a very distant one, between Ngan'gi and Murrinh-Patha, was based on the lexical data; Ngan'gi and Murrinh-patha have at most an 11% shared vocabulary density.

Present research is however overturning this view. Green (2003) has made a compelling case for Ngan'gi and Murrinh-patha making up a genetic subgroup

now labeled 'Southern Daly.' The case is based primarily on formal correspondences in the core morphological sequences of finite verbs. Green argues that these sequences match too closely in their complexities and irregularities to have plausibly come about through anything other than a shared genetic legacy; he demonstrates through reconstruction of finite verb paradigms that they are systematically derivable from an innovative common parent. The intriguing question, of how related neighboring languages have come to share as little as 11% lexical cognacy, remains unanswered.

Instead of a single 'Daly Family,' there now appears to be five separate Australian subgroups in the Daly region that cannot convincingly be related together as a single genetic unit (Table 1) (Green, 2003). Those similarities that Tryon took to be diagnostic of the 'Daly Family' are better accounted for either diffusionally or as genetically inherited features shared with a wide range of northern Australian languages.

Areal Features

Like the majority of languages spoken in Australia's central far north, Ngan'gi is of the polysynthetic (*see Central Siberian Yupik as a Polysynthetic Language*)

Table 1 Genetic subgroups in the Daly River region

| Subgroup | Principal language varieties |
|---------------|--|
| Anson Bay | Batytyamalh (aka Wadyiginy) (Wadjiginy) Kenderramalh (aka PunguPungu) |
| Northern Daly | MalakMalak (Mullúkumulluk), Tyeraty, Kuwema (Tyaraity) |
| Eastern Daly | Matngele |
| Western Daly | Kamu Marrithiyel (Marithiel), Marrisyefin, Marri Ammu Marringarr (Maringarr), Mati Ge Marramaninydyi (Marimanindji) Marranunggu (aka Warrgat) (Maranunggu), Emmi, Menhthe |
| Southern Daly | Murrinh-Patha Ngan'gikurunggurr (Nangikurunggurr), Ngen'giwumirri |

structural type, and is categorized within Australianist typology as belonging in the non-Pama-Nyungan and prefixing groups. It has complex verbal structures built up through the addition of strings of prefixes and suffixes to lexical roots. Some of the affixes are suppletive in form and many are portmanteau in nature, simultaneously encoding a number of grammatical categories. A large complex verb might have up to a dozen constituent morphemes and correspond in meaning to a whole English sentence. The most distinctive grammatical features of Ngan'gi are its extensive pronominal indexing, a set of 31 classifying verbs, 4 number categories for pronouns, a system of 16 noun classes, and a 3-way stop/fricative contrast.

Pronominal Indexing

Verbs obligatorily index core participants such as 'subject' (*I saw you*) and 'object' (*I saw you*) by bound pronominal prefixes. Most languages additionally allow for the verbal cross-referencing of other kinds of participants, such as 'goals' (*I told her the news*), 'benefactives' (*I cooked it for her*) and 'adversatives' (*My wife ran away on me*). Pronominal indexing is shown in (1), where subject, object, and adversative arguments are indexed on the one complex verb.

- (1) Danginy-nyi-fime-ngidde-wurru.
3sg.S.Poke.Perf-2sg.O-give-1sg.Adv-bad
'She gave it away to you against my wishes.'

Some grammatical categories, such as person, number, and tense, can be marked discontinuously via different affixes at different points in the verb. This marking is illustrated for subject number in (2), which stacks plural, dual, and trial affixes.

- (2) Ngarrgu-nime nge-rr-beny-gu-da-nime.
1exdlPRO-trial 1.S-nsg.S-Bash.Perf-dl.S-hit-tr.S
'We (trial exclusive) hit it.'

Verbal Classification

Many languages of Australia's central far north share the characteristic of forming their verbs with not one, but rather two, root-like elements. This two-part structure typically involves the pairing of a relatively inert root (or coverb), which provides the main lexical information for the verb, with a root that hosts the core grammatical affixes (or finite verb). Coverbs form an open class, while finite verbs constitute a small closed class – Ngan'gi has 31 of these. This two-part verbal structure is thought to be an ancient diffusional feature. While in some languages of Australia's north the finite verbs have synchronically no clear semantic value, in the Southern Daly languages, it functions as a classifier of the verbal

action. Verbal classification may simply involve specifying the relative orientation of the subject, as with the intransitive posture classifiers in (3) and (4).

- (3) Peke dini-fifi-tye.
tobacco 3sg.Sit.IMP-smoke-Past
'He was sitting smoking.'
- (4) Peke wirringe-fifi-tye.
tobacco 3sg.S.Sit.IMP-smoke-Past
'He was standing smoking.'

Other verbal classifiers are concerned with how objects are handled or manipulated, as illustrated by the transitive classifiers in (5) and (6). In (5), the finite verb *Hands* functions to conceptualize the action as performed within the grasp of the fingers. Contrast with (6), in which the replacement of *Hands* with *Poke* achieves a different schematic conceptualization, this time of the action as performed at the end of an elongated instrument.

- (5) Ngeriny-fityi peke.
1sg.S.Hands.Perf-roll tobacco
'I rolled a cigarette' (in my hands).
- (6) Ngariny-fityi screwdriver-ninggi.
1sg.S.Poke.Perf-roll screwdriver-INSTR
'I screwed it up with a screwdriver.' (I rolled it at the end of a long thin instrument).

When compared to the other languages of the Daly, and to northern Australia generally, the classifying verb structures of Ngan'gi reveal two aberrant features. First, they exhibit a tight morphophonological binding between coverb and the inflected finite root. Second, they show an innovative ordering, placing the coverb after the inflected finite root rather than preposed to it. Reid (2003) argued that these shared features result from recent diffusion rather than a shared genetic legacy, demonstrating how the Southern Daly languages Ngan'gikurunggurr and Ngen'giwumirri have acquired them only within the last hundred years.

Free Pronouns

The Daly languages have freeform pronoun systems that are complex by virtue of grammaticizing multiple nonsingular number categories. Some languages, including Ngan'gi, have singular/dual/trial/plural systems, while others have singular/dual/paucal/plural systems. As can be seen from Table 2, Ngan'gi has a slightly nonsymmetrical system where the trial/plural contrast is neutralized in 1st inclusive.

Nominal Classification

All the Daly languages have at least a few generic nouns, such as 'meat,' 'vegetable food,' and 'fire,' which are regularly placed in front of specific nouns

Table 2 Ngan'gi freeform pronouns

| Number | 1st inclusive | 1st exclusive | 2nd person | 3rd person |
|----------|-------------------|----------------------|---------------------|---|
| Singular | | <i>ngayi</i> | <i>nyinyi</i> | <i>nem</i> (male) <i>ngayim</i> (female) |
| Dual | <i>nayin</i> | <i>ngagarri</i> | <i>nagarri</i> | <i>wirrike</i> |
| Trial | <i>nayin nime</i> | <i>ngagarri nime</i> | <i>nagarri nime</i> | <i>wirrike nime</i> |
| Plural | <i>nayin nime</i> | <i>ngagurr</i> | <i>nagurr</i> | <i>wirrim</i> |

to encode salient cultural categories. In some Daly languages, this encoding has become extended into a system in which the category membership of all entities is obligatorily encoded by one of around a dozen NP initial generic nouns. In Ngan'gi, such generic-specific constructions have undergone even further grammaticalization, displaying agreement phenomena and reduction of the independent generic to bound forms. Sometimes, agreement is marked by bound forms attached to nouns as well as modifiers such as adjectives or demonstratives. In other cases, noun class assignment is marked by freeform generics that precede specific nouns and also precede the modifiers. Each of these types is demonstrated in (7) and (8).

- (7) a-tyalmerr a-kerre a-kinyi
animal-barramundi animal-big animal-this
 'this big barramundi fish'
- (8) syiri magulfu syiri marrgu
weapon fighting.stick weapon new
 'a new fighting stick'

Noun class phenomena in Daly languages have proved theoretically interesting by providing a perspective on the historical development of class markers from freeform nouns to proclitics to prefixes. They have also contributed to theorizing about the process by which agreement phenomenon develop (Reid, 1997) and to considerations of the nature of the distinction between noun class and noun classifying systems (Green, 1997).

Phonology

Australian languages generally lack phonemic fricatives and typically have just a single series of stop. The Daly region shows a significant departure from this pattern, with all languages except Anson Bay showing at least some phonemic voicing contrast. Ngan'gi has both a partial stop contrast, and phonemic fricatives, yielding a 3-way obstruent contrast between a voiced stop, voiceless stop and fricative for the bilabials and alveolars. The phonemes of Ngan'gi, showing the atypical obstruent set in an otherwise standard Australian inventory, are given in Table 3 and Table 4 in their practical orthography.

Table 3 Ngan'gi phonemic inventory consonants

| Consonant type | Bilabial | Alveolar | Palatal | Velar |
|----------------|----------|----------|---------|-------|
| Voiceless stop | p | t | ty | k |
| Voiced stop | b | d | | |
| Fricative | f | s | sy | g |
| Nasal | m | n | ny | ng |
| Lateral | | l | | |
| Approximant | | r | | |
| Trill | | rr | | |
| Glide | w | | y | |

Table 4 Ngan'gi phonemic inventory vowels

| Vowel type | Front | Back |
|------------|-------|------|
| High | i | u |
| Low | e | a |

Bibliography

- Dixon R M W (2002). *Australian languages: their nature and development*. Cambridge: Cambridge University Press.
- Evans N (2003). *Studies in language change 552: The non-Pama-Nyungan languages of Northern Australia: comparative studies in the continent's most linguistically complex region*. Pacific Linguistics: Canberra.
- Green I (1997). 'Nominal classification in Marrithiyel.' In Harvey M & Reid N (eds.) *Nominal classification in Aboriginal Australia*. Amsterdam: John Benjamins.
- Green I (2003). 'The genetic status of Murrinh-patha.' In Evans N (ed.) *Studies in Language Change 552: The non-Pama-Nyungan languages of Northern Australia: comparative studies in the continent's most linguistically complex region*. Pacific Linguistics: Canberra.
- Hoddinott W & Kofod F (1989). *The Ngankikurungkurr language (Daly River Area, Northern Territory)*. Series D, No. 77. Canberra: Pacific Linguistics.
- Reid N J (1990). *Ngan'gityemerri*. Ph.D. thesis: Australian National University: Canberra.
- Reid N J (1997). 'Class and Classifier in Ngan'gi.' In Harvey M & Reid N (eds.) *Nominal classification in Aboriginal Australia*. Amsterdam: John Benjamins.
- Reid N J (2000). 'Complex verb collocations in Ngan'gi: a non-derivational strategy for encoding valency alternations.' In Dixon R M W & Aikenvald A (eds.) *Changing valency*. Cambridge: Cambridge University Press.

- Reid N (2002). 'Sit right down the back: serialized posture verbs in Ngan'gi and other Northern Australian languages.' In Newman J (ed.) *Sitting, lying and standing: posture verbs in typological perspective*. Amsterdam: John Benjamins.
- Reid N J (2003). 'Phrasal verb to synthetic verb: recorded morpho-syntactic change in Ngan'gi.' In Evans N (ed.) *Studies in language change 552: The non-Pama-Nyungan languages of Northern Australia: comparative studies in the continent's most linguistically complex region*. Pacific Linguistics: Canberra.
- Reid N J (2005). *Ngan'gi dictionary*. Armidale: Australian Linguistics Press.
- Reid N & Harvey M (1997). *Nominal classification in Aboriginal Australian*. Amsterdam: John Benjamins.
- Tryon D (1974). *The Daly Family languages, Australia*. Series C, No. 32. Canberra: Pacific Linguistics.
- Walsh M J (1993). 'Languages and their status in Aboriginal Australia.' In Walsh M & Yallop C (eds.) *Language and culture in Aboriginal Australia*. Canberra: Aboriginal Studies Press.

Niger-Congo Languages

J Bendor-Samuel, Summer Institute of Linguistics, High Wycombe, UK

© 2006 Elsevier Ltd. All rights reserved.

It is widely accepted that the languages of Africa, apart from the creoles, pidgins, and 'official' languages inherited from the colonial era, belong to four major families: Niger-Congo, Nilo-Saharan, Afro-Asiatic, and Khoisan. Niger-Congo is the largest, both in terms of the number of languages and in geographical spread. It extends from Dakar east to Mombasa and south to Cape Town. Of the 2000 languages spoken in Africa, some 1400 belong to Niger-Congo. Recent estimates put the number speaking a Niger-Congo language at around 400 million people.

Early Classification

In the 19th century, scholars began to make groupings of African languages. Koelle (1854) published word lists in some 200 languages, grouped so as to reflect the relationships among the languages. Many of his groupings correspond closely to the accepted classification today.

Bleek (1856) recognized that languages in western and southern Africa were related and wrote of "that great family which, with the exception of the Hottentot dialects, includes the whole of southern Africa and most of the tongues of western Africa."

Subsequently, scholars tended to lose sight of the essential unity of these languages and to focus on the Bantu languages of southern Africa. The large number of languages elsewhere that had similar features were regarded as being 'mixed' in origin, and their similarities were explained as being the result of migrations and language contact, rather than as deriving from a common genetic origin with Bantu.

Significant development came with the work of Westermann (1927). He set up 'Western Sudanic' as distinct from 'Eastern Sudanic' (since classified as Nilo-Saharan). Westermann divided Western Sudanic into six subfamilies: Kwa, Benue-Congo, Togo Remnant, Gur, West Atlantic, and Mandingo (Maninka). He also compared a large number of proto-Western Sudanic roots with the corresponding proto-Bantu forms. Though Westermann did not go on to draw the conclusion that pointed to a common genetic origin, Joseph Greenberg did. He showed that Westermann's Western Sudanic and Bantu formed a single genetic family, which he called 'Niger-Congo.' Subsequently, Greenberg (1963) brought in Kordofanian as coordinate with Niger-Congo.

Greenberg retained the subfamilies that Westermann had already established – Kwa, Benue-Congo, Gur, West Atlantic, and Mande (Mandingo) – but included Togo Remnant within Kwa and added a new subfamily, which he termed 'Adamawa-Eastern.' His most revolutionary innovation was to include Bantu as a subgroup of a subgroup within Benue-Congo and not as a subfamily coordinate with the other main branches of Niger-Congo.

Accepted Classification

Although Greenberg's work has set the classificatory framework within which most scholars have worked since (Figure 1), there were, as he readily admitted, still many unresolved classificatory questions. Subsequent research has clarified some of these issues, and Greenberg's classification was revised by a group of scholars (Bendor-Samuel, 1989) (Figure 2). The languages that are spoken today are classified into nine major language subfamilies: Mande, Kordofanian, Atlantic, Ijoid, Kru, Gur, Adamawa-Ubangi, Kwa, and Benue-Congo. Scholars are not agreed on the classification of Dogon; hence,

it is listed separately, though it does not constitute a subfamily.

These nine major language subfamilies relate to each other in different ways, some being related more closely than others. The relationships reflect the fact that the nine major subfamilies did not derive directly from a common ancestor. There were intermediate steps that have been tentatively reconstructed on the chart.

Since the publication of *The Niger-Congo languages* (Bendor-Samuel, 1989), further consideration has been given particularly to three issues:

1. Although it is generally accepted that the five subfamilies of Kru, Kwa, Benue-Congo, Gur, and Adamawa-Ubangi are related more closely to each other than to Mande, Kordofanian, Atlantic, and Ijoid, there is no agreement on the sequence in which Mande, Kordofanian, Atlantic, and Ijoid split from the main stock. Some have suggested Kordofanian was the first to split off; others have proposed Mande.
2. As regards the Volta-Congo group, one view gaining acceptance is that the five subfamilies fall into two groups: Gur, Adamawa-Ubangi, and Kru

in North Volta-Congo and Kwa and Benue-Congo in South Volta-Congo. It has been suggested that Gur and Adamawa-Ubangi originated as a dialect continuum. Kwa and Benue-Congo could be treated in the same way. Although geographically separated from Gur, Kru is related more closely to Gur than to any of the other languages in Volta-Congo and so is placed in the Northern group. The attraction of this analysis is that it provides for a northern and southern spreading of languages in the Volta-Congo group, with Gur and Adamawa-Ubangi spreading across the Savannah lands to the north of this area, in contrast to Kwa and Benue-Congo spreading to the south (see Williamson and Blench, 2000).

3. Discussion has continued about where to draw the boundary between Kwa and Benue-Congo. Some have followed Greenberg in regarding the western groups of languages in Benue-Congo as an eastern group within Kwa. Others draw the boundary farther to the west, with these languages grouped within Benue-Congo, as is the case in this article (Williamson and Blench, 2000).

One thing is clear from the continuing discussion: the classification of Niger-Congo is far from final. Many questions are still being asked about the internal structures of the present subfamilies, their relationship to each other within Niger-Congo, and the relationship of Niger-Congo itself to Nilo-Saharan (Williamson and Blench, 2000).

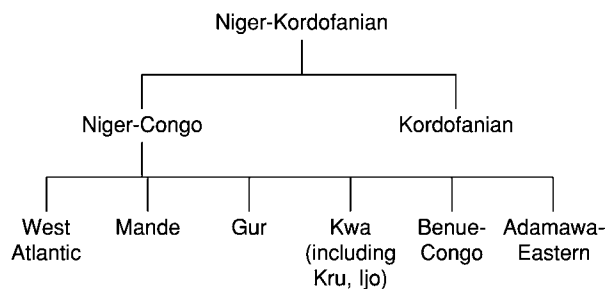


Figure 1 Greenberg's classification.

The Niger-Congo Subfamilies

Mande

Mande languages are spoken by over 10 million people over a wide area, including large parts of

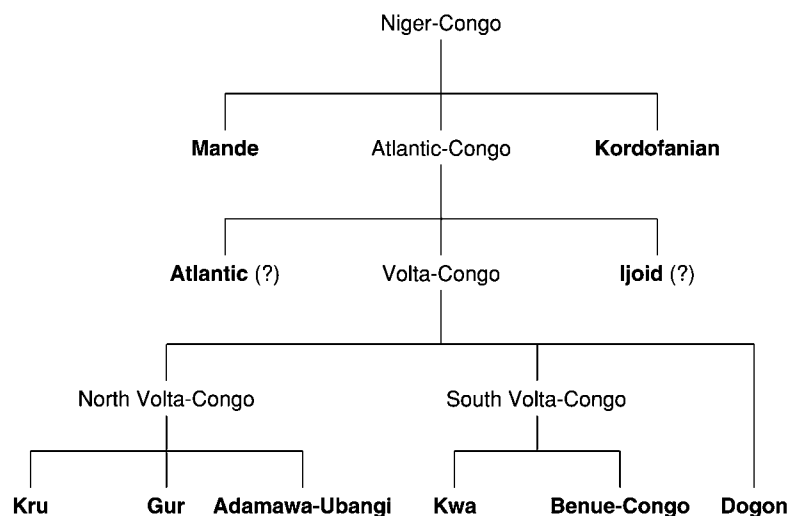


Figure 2 Revised classification.

Guinea, Mali, Sierra Leone, Liberia, and northwest Ivory Coast. Substantial numbers are also found in Burkina Faso, Senegal, Gambia, and Guinea-Bissau, with much smaller pockets in southern Mauritania; northern parts of Ghana, Togo, Benin, and Nigeria; and in southwest Niger.

The internal relationships of the languages within the Mande subfamily are loose. Lexicostatistical studies give only 17% cognates between the Western and Eastern groups.

The much larger Western group, with 27 languages, divides into a Southwestern group of 4 languages and a Northwestern group of 23 languages. Within the Northwestern group, the core group comprises 10 languages, all scoring from 80–90% lexicostatistically, with three of those languages – Bambara (Bamanakan), Maninka, and Jula – being major languages with over 1 million speakers each.

The 13 languages within the Eastern group relate to each other at levels of 30–35% lexicostatistically.

Kordofanian

Kordofanian languages are isolated from the rest of the languages of Niger-Congo, being spoken in the Nuba mountains of central Sudan by 250 000–500 000 people.

Kordofanian is divided into four main groups: Heiban, Talodi, Rashad, and Katla. Greenberg included a fifth group, Kadugli, but it is so divergent from the others that there is serious doubt whether it belongs to Kordofanian but rather to Nilo-Saharan (see Schadeberg, 1981).

Atlantic

Atlantic languages are spoken by about 20 million people. One language, Fula, accounts for 12–15 million of those people and is the most widely scattered language group in Africa, all the way from Senegal to the Sudan. Except for Fula, the Atlantic languages are located primarily along the Atlantic coast from the Senegal River to Liberia.

All of the Atlantic languages fall into one of two groups, northern and southern, except for the languages spoken on the Bijago Islands, which constitute a small third group with 20 000 speakers. Within the northern group, which includes Fula, Wolof, Serer, Jula (Jola), Manjaku-Papel, and Balanta (Balant-Ganja), lexicostatistical percentages range between 24–37.

The languages in the southern group are generally not related closely to each other. Three subgroups are recognized; Mel is the largest, comprising 13

languages, of which Temne (1 million speakers) and Bullom-Kissi (650 000) have the most speakers.

Ijoid

Ijoid is very different from the other subfamilies of Niger-Congo. It comprises the language cluster Ijo and a single language, Defaka. Ijo is not a single language, but a cluster of rather closely related languages/dialects with a total of over one million speakers. The whole subfamily is confined geographically to the Niger Delta. Ijoid does not belong either to Kwa or to Benue-Congo and seems to be outside the Volta-Congo grouping. Hence, it is treated as a branch of Atlantic-Congo.

Kru

Kru languages have been included by some scholars, including Greenberg, within Kwa, but later studies suggest that Kru is closer to Gur. Some have gone further and included Kru in a North Volta-Congo group together with Gur and Adamawa-Ubangi.

The 26 Kru languages are spoken by approximately 2 million people, mostly in the forest regions of southwest Ivory Coast and southern Liberia.

An Eastern and a Western group are recognized. The Western group is the larger and more heterogeneous and can be divided into four subgroups, each of which has one principal language complex: Grebo, Guere (Guere-Krahn), Bassa, and Klao. The Eastern group is more homogeneous and comprises two major subgroups, the Bete and the Dida language complexes, both in Ivory Coast.

Gur

The 85 Gur languages are found in the savannah lands north of the forest belt extending from southeast Mali across northern Ivory Coast, Burkina Faso, Ghana, Togo, and Benin into northwest Nigeria. The number of speakers of these languages is around 12–15 million.

Most Gur languages belong to Central Gur, which is a comparatively closely related group of languages. Within Central Gur, there are two major subgroupings, that of the Oti-Volta (some 25 languages) and Grusi (20 languages). Additionally, there are some 15 languages outside these two main subgroups. The languages within Oti-Volta show appreciably closer relationships to each other than do the languages within Grusi.

To the west of Central Gur lies the Senufo subgroup, comprising some 20 languages, grouped into 7 subgroups. Its relationship to Central Gur is not close, but there is no evidence to group Senufo with any other subfamily.

Dogon

Dogon is spoken in Mali, east of Mopti. Previously, scholars had included it within Gur, but there is general agreement that the grounds for this classification are inadequate. However, there is no evidence to group it with any of the other subfamilies within Volta-Congo.

Adamawa-Ubangi

Adamawa-Ubangi languages are spoken by approximately 8–9 million people from eastern Nigeria across northern Cameroon, southern Chad, the Central African Republic, and northern Zaire into southwestern Sudan.

Some lexicostatistical studies suggest that the Adamawa-Ubangi languages are closer to some of the Gur languages than they are to any of the languages in the other subfamilies within Volta-Congo. A preliminary hypothesis groups Gur and Adamawa-Ubangi in a North Volta-Congo grouping.

The languages within the Adamawa group are found mostly in Nigeria and Cameroon and are rather loosely related. The 70 plus languages/dialects are divided into 16 groups.

The Ubangi group has a larger number of speakers. The languages in it are related more closely to each other and include a number of widely spoken languages, such as Banda, Ngbandi, Ngbaka (Ngbaka-Mba), Gbaya, and Zande. The most probable classification suggests a core group comprising the three subgroups Banda, Ngbandi, and Ngbaka, with two peripheral groups Gbaya and Zande.

Kwa

The 45 Kwa languages stretch across southern Ivory Coast, Ghana, Togo, Benin, and into the southwest

corner of Nigeria, with a total of at least 20 million speakers.

This subfamily divides broadly into two main groups. The larger group, Nyo, comprises some 24 language clusters/languages/dialects covering most of southern Ghana and southeastern Côte d'Ivoire. Within Nyo, Potou-Tano is the largest subgroup with some 17 languages/dialects, including the major languages Akan (Twi and Fanti), Baule (Baoule), and the Guang cluster.

The smaller group, sometimes termed 'left bank' because its speakers live east of the Volta river, comprises seven language clusters/languages, of which the Gbe cluster (better known by the name of its largest member, Ewe) is the largest.

Benue-Congo

Benue-Congo is the largest of the subfamilies within Niger-Congo in terms of the number of languages, speakers, and geographical extent. It stretches from the Benin-Nigeria border across Nigeria eastward to Kenya and southward to the Cape. Thus, it covers over half the habitable terrain of the continent and a similar percentage of the population.

Benue-Congo is divided into 11 groups that can be arranged on an approximately west-to-east basis as in **Figure 3**.

All these groups, with the exception of Bantoid, are found primarily in Nigeria. The principal languages of each group are as follows: Defoid: Yoruba and Igala; Edoid: Edo and Urhobo; Nupoid: Nupe, Ibira (Ebira), and Gwari (Gbagyi); Idomoid: Idoma and Igede; Igboid: Igbo; Cross River: Efik, Ibibio, and Ogoni; Kainji: Kambari; Platoid: Berom, Tarok, and Jukun.

The 11th group, Bantoid, is the largest group in Niger-Congo, comprising several hundred languages,

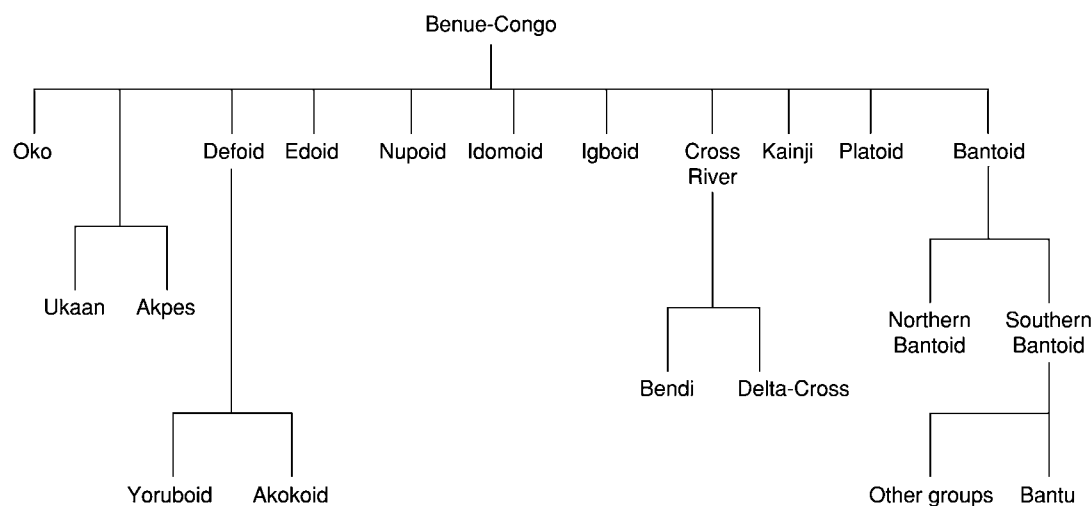


Figure 3 Benue-Congo subfamily.

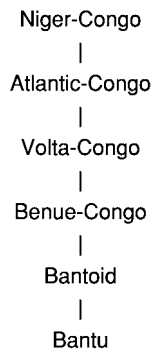


Figure 4 Bantu's genetic relationships.

covering most of the area southeast of Nigeria and Chad.

Bantoid is divided into a small northern group of languages spoken in eastern Nigeria and western Cameroon, and the very much larger southern group, which includes all the Bantu languages. Bantu's genetic relationships are illustrated in **Figure 4** (see **Bantu Languages**).

Bibliography

- Bendor-Samuel J (ed.) (1989). *The Niger-Congo languages*. Lanham, MD: University Press of America.
 Bleek W H I (1856). *The languages of Mozambique*. London: Harrison and Sons.

- Cole D T (1978). 'The history of African linguistics to 1945.' In Sebeok T A (ed.) *Current trends in linguistics, 7: Linguistics in sub-Saharan Africa*. The Hague & Paris: Mouton. 1–29.
 Greenberg J H (1963). *The languages of Africa*. The Hague: Mouton.
 Koelle S W (1854). *Polyglotta Africana*. London: Church Missionary House.
 Schachter P (1971). 'The present state of African linguistics.' In Sebeok (ed.), 30–44.
 Schadeberg T C (1981). 'The classification of the Kadugli language group.' In Schadeberg T C & Bender M L (eds.) *Nilo-Saharan*. Dordrecht. Foris Publications. 291–305.
 Sebeok T A (ed.) (1971). *Current Trends in Linguistics 7: Linguistics in Sub-Saharan Africa*. The Hague: Mouton.
 Watters J R (1989). 'Bantoid overview.' In Bendor-Samuel (ed.), 401–420.
 Westermann D (1927). *Die westlichen Sudansprachen und ihre Beziehungen zum Bantu*. Berlin: de Gruyter.
 Westermann D & Bryan M A (1952). *Languages of West Africa: Handbook of African languages*, vol. 2. London: Oxford University Press.
 Williamson K (1989). 'Niger-Congo overview.' In Bendor-Samuel J (ed.), 3–45.
 Williamson K (1989). 'Benue-Congo overview.' In Bendor-Samuel J (ed.), 247–274.
 Williamson K & Blench R (2000). 'Niger-Congo.' In Heine H & Nurse D (eds.) *African languages, an introduction*. Cambridge: Cambridge University Press. 11–42.

Nilo-Saharan Languages

G J Dimmendaal, University of Cologne, Cologne, Germany

© 2006 Elsevier Ltd. All rights reserved.

One of Africa's major language phyla, Nilo-Saharan consists of at least 120 languages spoken in an area covering major areas in eastern and central Africa, with a westward extension as far as the Niger Valley in Mali, West Africa. The genetic unity of these languages was first proposed by Greenberg (1963) on the basis of recurring morphological features and lexical similarities. According to Greenberg, Nilo-Saharan constitutes one of the four phyla on the continent, next to Afroasiatic, Khoisan, and Niger-Congo.

Greenberg initiated his classificatory work on African languages in the late 1940s and early 1950s when he established, among others, a Macro-Sudanic family, consisting of Eastern Sudanic, Central Sudanic, Berta, and Kunama (cf. Greenberg, 1955, which contains a collection of articles published earlier in the

Southwestern Journal of Anthropology). Macro-Sudanic was subsequently renamed Chari-Nile, after the two major rivers in the area. Based on a judicious evaluation of the available data, Greenberg (1963) pulled together several disparate groups formerly considered linguistic isolates and known primarily through the pioneering work of Tucker and Bryan (1956) into a new phylum called Nilo-Saharan, of which Chari-Nile formed the core, with Songai (Songhay), Saharan, Maban plus Mimi, For (or Fur), and Koman as additional primary branches (see **Table 1**).

There are at least 120 distinct Nilo-Saharan languages, the number of speakers for individual languages ranging from several millions (compare the sections on Dinka, Kanuri, Luo, and the Songai cluster) to languages with only a few speakers, e.g., Aka, Kelo, and Molo, which belong to the Jebel group within Eastern Sudanic. Although considerable progress has been made over the past few decades with the description and comparison of several Nilo-Saharan

Table 1 Nilo-Saharan subgroups

| <i>Greenberg, 1955</i> | | <i>Greenberg, 1963, 1971</i> | <i>Current nomenclature</i> |
|------------------------|---------|------------------------------|-----------------------------|
| Songhai | isolate | Songhai | Songai |
| Central Saharan | isolate | Saharan | Saharan |
| Maban | isolate | Maban | Maban |
| Mimi | isolate | Mimi | Mimi |
| Fur | isolate | Fur | For |
| Nyangiyan | isolate | | Kuliak (also: Rub) |
| | | | Temein, Keiga Jirru |
| Temainian | isolate | | Nubian |
| Nubian | | | Surmic |
| Beir-Didinga | | | |
| Barea | | | Nara |
| Tabi | | | (West) Jebel |
| Nyimang | | | Nyimang, Dinik |
| Merarit | | | Taman |
| Dagu | | | Daju |
| Nilotic | | | Nilotic |
| Great Lakes | | | |
| Central Sudanic | | | Central Sudanic |
| Berta | | | Berta |
| Kunama | | | Kunama |
| Koman | isolate | | Koman |

subgroups, a number of lower-level units, for example Daju or Koman, remain poorly known.

Several Nilo-Saharan languages are used only in oral communication; for others orthographies in Latin, Arabic, or Fidäl script have been developed. Old Nubian, which was written in a modified Coptic script, dates back to the 8th century of the Christian era. The role of Nilo-Saharan languages in education also varies considerably depending on the number of speakers, and also on the language policy in the countries where these languages are spoken. For major languages such as Kanuri or Luo, but also for other languages, in particular those spoken in Kenya and Uganda, there is a growing body of literature containing novels, poetry, and oral traditions. Apart from written texts and reference grammars, dictionaries have become available more recently for several Nilo-Saharan languages, for example, Keegan (1996) on the Central Sudanic language Mbay, Creider and Creider (2001) on Nilotic Nandi, or Heine (1999) on the Kuliak (Rub) language Ik.

More Recent Comparative Work

To date, neither the limits of Nilo-Saharan nor the internal organization has been settled. This situation

is due to heterogeneity within the phylum (with several genetically isolated languages) and the presumably considerable time depth involved, as well as the paucity of descriptive sources for a variety of languages. The genetic status of the Songai cluster on the great bend of the Niger River, mainly in Mali and Niger and extending into neighboring countries, remains debatable, for example. However, as a result of the pioneering descriptive work on several varieties of Songai by Heath (e.g. 1999a), future in-depth comparative work with respect to this cluster may be expected.

Some progress has been made over the past decades with the historical comparison of lower-level units such as Central Sudanic, Daju, Koman, Maban, Nilotic, Nubian, Saharan, and Surmic. More recently, it has been argued by Rilly (2003) that the extinct language of the Meroitic empire, preserved in written records which have only been partly deciphered, not only shows Eastern Sudanic affinity, as already proposed by Greenberg (1971), but that it was most closely related to Eastern Sudanic groups such as Nubian, Taman, Nara, and Nyimang (plus Dinik).

On the basis of an extensive comparison of lexical entries, presumed sound correspondences, and grammatical comparison, Ehret (2001) has regrouped various Nilo-Saharan units; according to this classification using shared phonological and grammatical innovations for subclassification, Central Sudanic and Koman, which are also typologically rather distinct from remaining Nilo-Saharan groups, constitute genetic outlayers.

Bender (1996, 2000) has also proposed lexical isoglosses and grammatical isomorphs for Nilo-Saharan. Moreover, the same author assumes that Songai, Saharan, and Kuliak constitute primary branches of Nilo-Saharan, whereas the remaining subgroups form a fourth branch. Unlike Ehret, Bender assumes that the Koman group is most closely related to the Eastern Sudanic group. Moreover, he assumes, as other scholars have done, that the Kadu languages in the Nuba Mountains are also part of Nilo-Saharan.

Controversy also remains over the inclusion or exclusion of languages like Biraile (Birale; also known as Ongota), whose speakers live along the Weyt'o River in Ethiopia, and the Shabo (or Mekeyir), another small ethnic group living in southwestern Ethiopia.

The Areal Dimension

Corresponding to the wide geographical spread of Nilo-Saharan languages, considerable typological diversity exists between them. A number of properties nevertheless are widespread and in fact are shared with neighboring Niger-Congo, and to a lesser extent

Afroasiatic, languages. For example, ATR-vowel harmony, in its classical form involving a set of [–advanced tongue root] vowels (ɪ, ɛ, a, ɔ, ʊ), and a set of [+advanced tongue root] vowels (i, e, ä, o, u), is attested in varieties of Songai, For, Kunama, Eastern Sudanic groups like Nubian, Temein, Nilotic, and Surmic as well as in Koman languages. The role of areal contact in this respect remains to be determined. Less common are systems with seven vowels (e.g., in the Nilotic language Datooga) or five vowels (e.g., in Nara). As shown by Andersen (1991), the contrast between breathy and creaky voice vowels in the Nilotic language Dinka goes back historically to an ATR contrast.

Another areal property of Nilo-Saharan shared with neighboring Niger-Congo and Afroasiatic languages is tone, with systems varying between classical two-tone systems with downdrift and downstep and systems with up to four distinct level tones which may also form complex (contour) tones. A number of Nilo-Saharan languages spoken along the northern edge and bordering on Afroasiatic, such as the Songai language Koyra Chiini, appear to be nontonal (Heath, 1999b). There are relatively few tonal grammars of Nilo-Saharan languages. Also, the historical-comparative study of tonal systems is still in its initial stage, Boyeldieu (2000) on a group of Central Sudanic languages being one of the few modern studies in this respect.

Consonant systems range from fairly simple, e.g., thirteen consonants in Southern Nilotic Kalenjin, to a wide range of contrasts in Central Sudanic and Koman languages. Here too, areal contact appears to have played a role; the Kalenjin consonant system, for example, is similar to that of neighboring Bantu (Niger-Congo) languages. Central Sudanic languages like Ngiti have a contrast between voiced and voiceless implosive stops, according to Kutsch Lojenga (1994); moreover, words in Ngiti as well as in the closely related language Lendu may consist of syllabic consonants like *s*, *z*, or *r* only. Whereas voiced implosive stops are more common across Nilo-Saharan, such stops are found in combination with ejectives in Berta, Koman and Surmic; a similar contrast is found in neighboring Omotic (Afroasiatic) languages. The dental/alveolar contrast for stops appears to be restricted to Eastern Sudanic groups like Nilotic, Surmic, Temein, Central Sudanic Kreish, or the Koman language Kwanimpa. Labial velars stops are common in Central Sudanic as well as neighboring Nilotic languages. The role of areal contact with Niger-Congo languages again remains to be determined.

A prototypical feature of many Nilo-Saharan groups, which appears to be relatively rare elsewhere

in the world (except in Cushitic and Semitic, i.e., Afroasiatic, languages), involves a distinction between singulative, plural or collective, and replacement marking (Dimmendaal 2000). Nouns referring to items usually occurring in pairs or in larger numbers, such as ‘breast’, ‘fly’, or ethnonyms, tend to take a singulative marker and are morphologically unmarked in the plural. In addition to plural (or collective) marking, nouns (in particular derived ones) tend to be inflected for number both in the singular and the plural. This three-way distinct is not attested in geographically more peripheral Nilo-Saharan zones, e.g., in Central Sudanic, Saharan, or Songai. Gender marking, though not widespread, is found as an inflectional feature of nouns in Eastern Nilotic, for example, and as a derivational property in the Southern and Western branch of Nilotic.

From a morphosyntactic point of view, Nilo-Saharan language groups spoken in an area ranging from northern Ethiopia and Eritrea across north-central Sudan and extending into Chad and Nigeria share typological features with Afroasiatic languages in Ethiopia. These include a basic constituent order whereby the verb occurs in final position, an extensive case marking system, verbal compounding (e.g., with ‘say’ or other types of light verbs, such as ‘put’ or ‘do’), as well as the use of converbs as dependent verb forms in complex sentences, although not all properties are necessarily present in all groups. These common typological features to some extent may be due to areal diffusion as a result of long-term cultural contacts and corresponding patterns of multilingualism between speech communities in these areas. The Wadi Howar, also known as the Yellow Nile (a former river sanctuary and tributary to the Nile which connected the mountainous area in eastern Chad with the Nile Valley from about 8000 B.C. till about 1000 B.C.; cf. **Figure 1**), possibly constituted an important geographical condition for this cultural and linguistic diffusion (cf. Keding, 2000, for a discussion of the geomorphological and archaeological background; see also Amha and Dimmendaal (2006), for a discussion). The gradual extinction of this riverine system may have resulted in a diaspora of Nilo-Saharan languages from the Wadi Howar region in an eastern, western, and southern direction.

Ehret (2001: 202–209) has argued for the reconstruction of a series of case suffixes for the earliest stages of Nilo-Saharan. Such dependent-marking systems are indeed attested in a range of Nilo-Saharan groups between Chad and Eritrea, as pointed out above. Remnants of case marking are also attested in Central Sudanic languages, but not apparently in Koman or Songai. Reduced case marking systems, and, correspondingly, a more extensive verbal

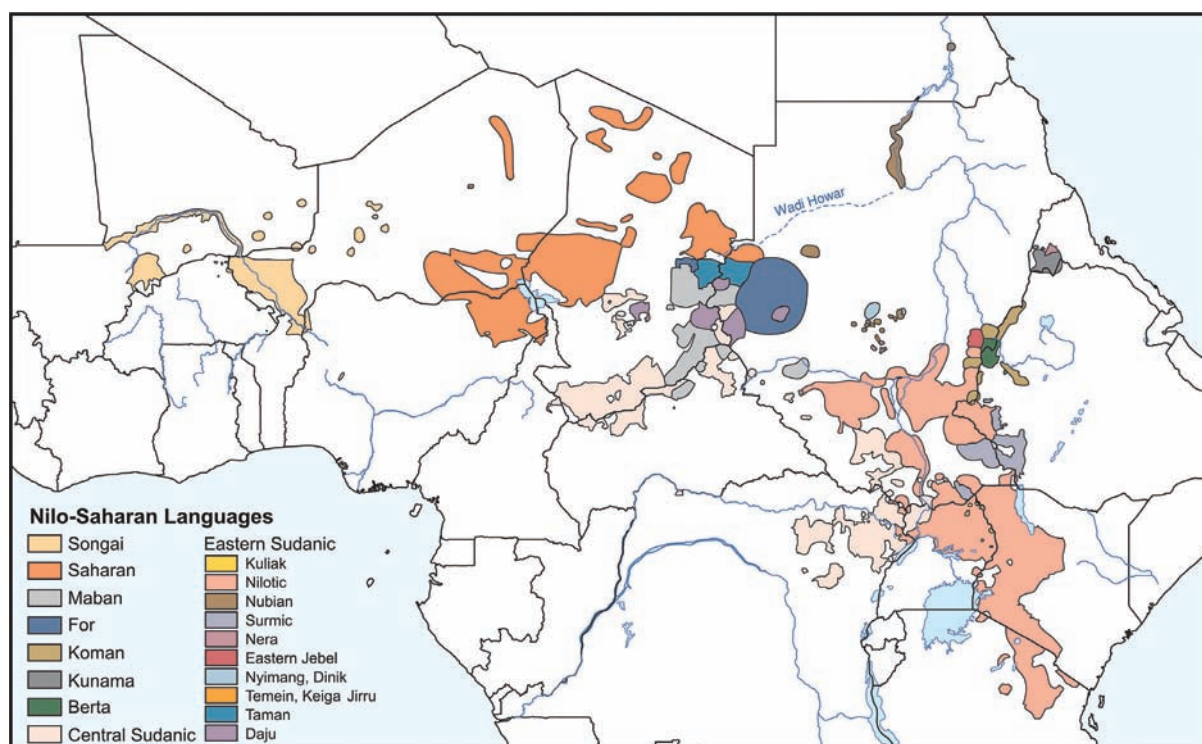


Figure 1 Map of Nilo-Saharan languages.

strategy of marking semantic roles like location, direction, and instrument, is found in Berta and Eastern Sudanic groups like Daju, Nilotic, Surmic, and Temein. Rather than having preverbal subjects, several languages belonging to these latter groups allow for postverbal subjects which are marked for case (the so-called ‘marked nominative’). An interesting combination of head marking and dependent marking at the clausal level is found in the verb-initial Kuliak (Rub) languages (cf. König, 2002). For a number of Nilotic languages using the marked nominative strategy for postverbal subjects, it has been argued that this applies only to transitive predications, thus giving rise to ergative properties, as the object or the subject of an intransitive predicate precede the verb and are not marked for case. As shown by Reh (1996) for the Nilotic language Anywa (Anu), OVS and SV are but one of several constituent order types allowed for in the language, alternatives being governed by pragmatic principles (e.g. active and less active participants, or participant orientation as against action orientation).

Pronominal subject (and occasionally object) marking, diathesis, causative, and pluractional marking are common morphological properties of verbs in Nilo-Saharan. Languages in the border area between Sudan and Chad, such as For or the Maban and Taman group, manifest complex morphophonemic alternations for consonants in their verb systems.

Derivational morphology tends to be expressed mainly by way of suffixation, and by prefixation in Central Sudanic; the causative marker, however, involving a presumably cognate morpheme consisting of a high front vowel, tends to be expressed as a prefix in Central Sudanic as well as elsewhere in Nilo-Saharan.

The use of logophoric pronouns marking coreferentiality between argument positions across clauses and sentences is a feature of some Nilo-Saharan groups which is shared with Niger-Congo and Chadic as well as Omotic (i.e. Afroasiatic) languages. A universally uncommon anti-logophoricity marking is found in Western Nilotic Mabaan.

Bibliography

- Amha A & Dimmendaal G J (2006). ‘Converbs from an African perspective.’ In Ameka F, Dench A & Evans N (eds.) *Catching Language*. Berlin and New York: Mouton de Gruyter.
- Andersen T (1991). ‘Vowel length in Western Nilotic Languages.’ *Acta Linguistica Hafniensia* 22, 5–26.
- Bender M L (1996). *LINCOM handbooks in Linguistics 6: The Nilo-Saharan languages*. Munich: LINCOM Europa.
- Bender M L (2000). ‘Nilo-Saharan.’ In Heine B & Nurse D (eds.) *African languages: an introduction*. Cambridge: Cambridge University Press. 43–73.
- Boyardieu P (2000). *Sprache und Geschichte in Afrika 10: Identité tonale et filiation des langues sara-bongo-baguiriennes (Afrique Centrale)*. Cologne: Rüdiger Köppe.

- Creider J, Tapsubei & Creider C A (2001). *Nilo-Saharan linguistic analyses and documentation 16: A dictionary of the Nandi language*. Cologne: Rüdiger Köppe.
- Dimmendaal G J (2000). 'Number marking and noun categorization in Nilo-Saharan languages.' *Anthropological Linguistics* 42, 214–261.
- Ehret C (2001). *Sprache und Geschichte in Afrika 12: A historical-comparative reconstruction of Nilo-Saharan*. Cologne: Rüdiger Köppe.
- Greenberg J H (1955). *Studies in African linguistic classification*. New Haven: Compass.
- Greenberg J H (1963). 'The languages of Africa.' *International Journal of American Linguistics* 29(1). Part 2.
- Greenberg J H (1971). 'Nilo-Saharan and Meroitic.' In Sebeok T A (ed.) *Current trends in linguistics vol. 7: Linguistics in sub-Saharan Africa*. The Hague: Mouton. 421–442.
- Heath J (1999a). *A grammar of Koyra Chiini: the Songhay of Timbuktu*. Berlin and New York: Mouton de Gruyter.
- Heath J (1999b). *Westafrikanische Studien 19: A grammar of Koyraboro (Koroboro) Senni: the Songhay of Gao, Mali*. Cologne: Rüdiger Köppe.
- Heine B (1999). *Nilo-Saharan linguistic analyses and documentation 15: Ik dictionary*. Cologne: Rüdiger Köppe.
- Keding B (2000). 'New data on the Holocene occupation of the Wadi Howar region (Eastern Sahara/Sudan).' *Studies in African Archaeology* 7, 89–104. [Poznań Archaeological Museum].
- Keegan J M (1996). *Dictionary of Mbay*. Munich: LINCOM Europa.
- König C (2002). *Nilo-Saharan linguistic analyses and documentation 17: Kasus im Ik*. Cologne: Rüdiger Köppe.
- Kutsch Lojenga C (1994). *Nilo-Saharan linguistic analyses and documentation 9: Ngiti: a Central-Sudanic language of Zaire*. Cologne: Rüdiger Köppe.
- Reh M (1996). *Nilo-Saharan linguistic analyses and documentation 11: Anywa language: description and internal reconstructions*. Cologne: Rüdiger Köppe.
- Rilly C (2003). *Le méroïtique: données grammaticales, lexicales. Position linguistique. Thèse de doctorat, Ecole Pratique des Hautes Etudes, Paris*.
- Tucker A N & Bryan M A (1956). *The non-Bantu languages of north-eastern Africa*. London: Oxford University Press for the International African Institute.
- Tucker A N & Bryan M A (1966). *Linguistic analyses: the non-Bantu languages of north-eastern Africa*. London: Oxford University Press for the International African Institute.

Niuean

W B Sperlich, Auckland, New Zealand

© 2006 Elsevier Ltd. All rights reserved.

Niuean (Niue) is the language of Niue, a Pacific island and self-governing territory of New Zealand. Niuean belongs to the Tongic subgroup of Polynesian (with Tongan and Niuean as the only members). The language name is synonymous with the name of the island. It is derived from *niu-ee* ('coconut-see') and is considered by oral tradition to be the exclamation by the earliest arrivals from Tonga (no earlier than 2000 years ago, according to archaeological evidence), who were surprised to see many coconut palms growing on the island. While Niuean is clearly a Tongic language (i.e., the first and subsequent settlers arrived mainly from Tonga) there are also elements of Samoan, Pukapuka, and Cook Islands Maori embedded in the Niuean language. Borrowings since European contact, especially early on via missionary efforts and more recently via trade, tourism, and globalization, are on the increase, and the impact of English syntax on Niuean is already pronounced. However, conservative Niuean is a strict verb-initial (or predicate-initial) language, and depending on the definition of 'subject' and 'object' can be designated

as either VSO or VOS. It has been argued that subjecthood in Niuean is not a syntactic category and that only a core predicate and arguments are the basic constituents of a sentence. Niuean is a split-ergative language (morphologically ergative, syntactically accusative). A canonical sentence with ergative (ERG) and absolutive (ABS) case marking is exemplified below:

| | | | | | | |
|----|-----|-----|------|------|-----|---------|
| Ne | kai | he | pusi | ia | e | moa |
| T | eat | ERG | cat | that | ABS | chicken |

'that cat ate the chicken'

To better express the ergative case above, a more literal (but less idiomatic) translation would be via the English passive: 'The chicken was eaten by that cat.' Indeed, some linguists have compared the ergative case marking to that of passives, noting, however, that the ergative case is unmarked. Niuean does not have antipassives. Other syntactic areas of special interest include raising, instrumental advancement, topicalization, causatives, possession, reflexives, and noun incorporation. A special feature of Niuean morphology is reduplication. Furthermore, highly productive affixation allows complex morphological strings. The causative prefix *faka-* and the prefix *ma-*, which changes verbs into participles, are most

common. Various degrees of lexicalization can make analysis difficult. The phonological inventory is simple in its use of only some 10 consonants (and no clusters) but is complex in its use of vowels, which are either short or long. Short vowels give rise to practically all combinations of diphthongs. Since the syllable structure is (C)V(V), there are many extended vowel sequences across morpheme boundaries, as for example in the complex word *faka-fe-haga-ao-aki*, where the sequence *-alaola-* is rearticulated.

Niuean as a conservative language is an important witness for historical linguists who work on proto-Polynesian and proto-Oceanic languages, as well as being an important sample language in typological and comparative linguistics.

Bibliography

- Kirikiri R (1974). Aspects of the case for Niuean. M.A. thesis, University of Auckland.
- Massam D & Levin J (1988). 'Raising and binding in Niuean.' In McGinn D (ed.) *Studies in Austronesian linguistics*. Athens: Ohio University. 253–274.
- Massam D & Roberge Y (1997). 'Operator-bound clitics and Niuean ai.' In Black J & Motapanyane V (eds.) *Clitics, pronouns and movement, current issues in linguistic theory*. Amsterdam: Benjamins. 273–300.
- Massam D (1998). 'Instrumental aki and the nature of Niuean transitivity.' *Oceanic Linguistics* 37(1), 12–28.
- Massam D (2000). 'VSO is VOS: aspects of Niuean word order.' In Carnie A & Guilfoyle E (eds.) *The syntax of verb initial languages*. Oxford: Oxford University Press. 97–117.
- Massam D & Sperlich W (2000). 'Possession in Niuean.' *Language Typology and Universals (Sprachtypologie und Universalienforschung)*. Special Volume on Possessive Markers in Central Pacific Languages, 281–292.
- Massam D (2001). 'On predication and the status of subjects in Niuean.' In Davies W & Dubinsky S (eds.) *Objects and other subjects*. Amsterdam: Kluwer. 225–246.
- Massam D (2001). 'Pseudo noun incorporation in Niuean.' *Natural Language and Linguistic Theory* 19(1), 153–197.
- Massam D, Abdolhosseini M H & Oda K (2002). 'Number and events: verbal reduplication in Niuean.' *Oceanic Linguistics* 41(2), 475–492.
- Massam D (2002). 'Some notes on the quantifier *oti* in Niuean.' *Rongorongo Studies: A Forum for Studies in Polynesian Philology* 12(2), 56–65.
- Massam D (2004). 'Predicate fronting and lexical category in Niuean.' In Carnie A, Harley H & Dooley-Colberg S (eds.) *Verb first: studies in predicate initial languages*. Amsterdam: Benjamins. 227–242.
- McEwen J M (1970). *Niue Dictionary*. Wellington: Department of Maori and Island Affairs.
- Seiter W (1980). *Studies in Niuean syntax*. New York: Garland Press.
- Sperlich W B (1996). 'Niuean syntax under pressure from English.' *Rongorongo Studies: A Forum for Studies in Polynesian Philology* 6(1), 17–27.
- Sperlich W B (1997). *Niue language dictionary*. Hawai'i: University of Hawai'i Press.
- Sperlich W B (2000). 'Is Niuean a 100% Tongic language?' In Fischer S & Sperlich W B (eds.) *Leo Pasifika: Proceedings of the Fourth Conference on Oceanic Languages*. Auckland: Institute of Polynesian Languages and Literatures. 343–353.
- Sperlich W B (2001). 'Semantic and syntactic functions of reduplication in Niuean.' In Bradshaw J & Regh K L (eds.) *Issues in Austronesian Morphology*. Pacific Linguistics 519: The Australian National University.
- Sperlich W B (2002). 'Inside and outside Niuean space.' In Bennardo G (ed.) *Pacific Linguistics 523: Representing space in Oceania*. Canberra: The Australian National University. 77–90.
- Sperlich W B (2004). 'Borrowing in Niuean.' In Tent J & Geraghty P (eds.) *Pacific Linguistics 548: Borrowing, a Pacific perspective*. Canberra: The Australian National University. 295–306.
- Sperlich W B (2005). 'Can cyber-chat save an endangered language: A Niue case study.' *Journal of Sociolinguistics* 9.
- Whittaker G (1982). *The Niuean language: an elementary grammar and basic vocabulary*. Alofi: University of the South Pacific, Niue Centre.
- Wilson W H (1989). 'A-O possessive neutralization in Niuean.' In VICAL 1, *Oceanic Language Papers from the Fifth International Conference on Austronesian Linguistics*. Wellington: Linguistic Society of New Zealand.

Nivkh

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

Nivkh (Gilyak) has perhaps 400 speakers (1991) out of an ethnic population of 4400 (as of 1996; G. A. Otaina) in East Siberia. There are two dialects,

the Amur and the Sakhalin; the latter is subdivided into eastern and northern clusters. There are approximately 100 Amur Nivkh speakers out of a population of 2000 and 300 Sakhalin Nivkh speakers out of a total population of 2700 (as of 1995; M. Krauss). On Sakhalin Island, many Nivkh live in the villages of Nekrasovka and Nogliki. Along the Amur River, a number of Nivkh reside in Aleevka village. Nivkh is

a true language isolate. Attempts to link it with other groups have never succeeded. The Nivkh still maintain their long-practiced traditional economies based on subsistence fishing supplemented by hunting and gathering.

Word initially, the velar nasal is common in Nivkh, as shown in Example (1) (Gruzdeva, 1998: 11–13, 24, 32):

- (1) *ɲaarla* ‘very fat’ *ɲax* ‘6’
 ɲax ‘6’ *ɲamk* ‘7’
 ɲiŋk ‘face’ *ɲaɲfɲaɲf* ‘each other’

One of the hallmarks of Nivkh structure is the curious and characteristic system of morphologically conditioned stem-initial consonant mutation to mark a range of inflectional (and derivational) categories. In this regard, note the initial consonant in the word for ‘head’ in the following three Nivkh forms (Gruzdeva, 1998: 14):

- | | | | |
|-----|---------------------------|----------------|------------------------|
| (2) | kyxkyx t ^h oŋr | it zoŋr | čam d ^h oŋr |
| | swan head | drake head | eagle head |
| | 'swan's head' | 'drake's head' | 'eagle's head' |

Like many languages of the Pacific Rim, Nivkh contrasts special portmanteau counting forms for nouns of various types (e.g., people vs. animals). Twenty six such classes of numerals have been reckoned for Nivkh by Panfilov (1962) and Krejnovich (1932). Compare the following examples for Amur and East Sakhalin dialects (Panfilov, 1962: 6–7; Gruzdeva, 1998: 24):

- | | | | |
|-----|-------------|--------------------|--------------|
| (3) | Amur | E. Sakhalin | Gloss |
| | ñin, ñen | ñenn | '1 person' |
| | nir | nirŋ | '4 people' |
| | ŋax | ŋax | '6 people' |
| | ŋamk | ŋamk | '7 people' |
| | Amur | E. Sakhalin | Gloss |
| | ñiñ | ñan | '1 animal' |
| | nur | nur | '4 animals' |
| | ŋax | ŋax | '6 animals' |
| | ŋamk | ŋamk | '7 animals' |

Nivkh is typical of Siberian languages in its use of a range of grammatical and local/directional case forms. Among the oppositions found in Nivkh is a contrast between a dative and an allative case, as shown in the following examples (Gruzdeva, 1998: 20, 21) (NEG, negative; DAT, dative; IMP, imperative; REFL, reflexive; LOC, locative; ABL, ablative; ALL, allative; TERM, terminative; FUT, future; FIN, finite; PRED, predicative):

- (4a) t^ha ikin-doχ t^haxta-ya
 NEG elder.brother-DAT be.angry-IMP
 ‘Don’t be angry at (your) elder brother.’

- (4b) nin-doʒ p^h-vo-x
 we-DAT REFL-village-LOC/ABL what
 t^hamdid xer-ya
 tell-IMP
 ‘Tell us what is (going on) in your village.’
- (4c) ni eri-rʒa vi-ni-d^j-ra
 I river-ALL/TERM go-FUT-FIN-PRED
 ‘I shall go (up) to the river.’
- (4d) tf-itk haimʒaf-toʒo hunv-nd-ra
 2-father old.age-ALL/TERM live-FIN-PRED
 ‘Your father lived (up) to old age.’

One category marked morphologically in the Nivkh verb is reciprocal action. This is encoded by the prefix *v-* in Amur Nivkh and *v-/o-/u-* in Sakhalin Nivkh.

- (5) v- or 'meet' v-ayay 'disturb each other'
o-smu 'love each other'

Also found is a syntactic means of indexing this category through the word *p^hnafqnafq* (REFL-friend. friend) ‘each other,’ a possible calque from Russian *drug druga*, etc.; the verbs in such sentences lack the reciprocal prefix (Gruzdeva, 1998: 32):

- (6) imŋ p^hŋafqŋafq. lov-di
 they each.other imitate-FIN
 'They imitated each other.'

Subordinate clauses in Nivkh are generally marked by some kind of nonfinite, nominalizing, or adverbializing morphology. This is common in many Siberian and other Eurasian languages (Anderson, 2003). The following example is from Gruzdeva (1998: 50):

- (7) imk čo hak-vul
mother fish cut-TEMP.CV
p^h-ajmɲaʁ-kir̥ roř k^herai-d
REFL-husband-INSTRUMENTAL together talk-FIN
'Mother talked with her husband while cutting
fish.'

Bibliography

- Anderson G D S (2003). 'Towards a phonological typology of native Siberia.' In Holisky D A & Tuite K (eds.) *Current trends in Caucasian, East European and Inner Asian linguistics. Papers in honor of Howard I. Aronson*. Amsterdam/Philadelphia: John Benjamins. 1–22.
- Austerlitz R (1982). 'Gilyak internal reconstruction, 1: seven etyma.' In Aronson H I & Darden B J (eds.) *Papers from the second conference on the Non-Slavic languages of the USSR. Folia Slavica 5(1–3)*. Columbus: Slavica. 81–88.
- Austerlitz R (1984). 'Gilyak internal reconstruction, 2: iron and questions related to metallurgy.' In Aronson H I & Darden B J (eds.) *Papers from the third conference on the Non-Slavic languages of the USSR. Folia Slavica 5(1–3)*. Columbus: Slavica. 93–102.
- Comrie B (1981). 'Nivkh.' In *Languages of the Soviet Union*. Cambridge: Cambridge University Press. 266–272.

- Gruzdeva E (1998). 'Nivkh.' *Languages of the World Materials LW/M 111*. München: Lincom-Europa.
- Hattori T (1962). 'Versuch einer Phonologie des Südostgiljakischen – Phonembestand, Verteilung und Alternation.' *Journal of Hokkaido Gakugei University (Sapporo), Section I-A 13*, 67–130. *Section II, 13*, 29–96. [Special reprint.]
- Jacobson R (1971). 'Notes on Gilyak.' In *Selected writings: II. Word and language*. The Hague, Paris: Mouton. 72–102.
- Krejnovich E A (1932). *Giljatskie chislitel'nye. Trudy Nauchno-issledovatel'skoj assotsiatsii* (tom I, vypusk 3). Leningrad: Instituta Narodov Severa.
- Krejnovich E A (1934). 'Nivxskij (giljatskij) jazyk.' In Krejnovich E A (ed.) *Jazyki i pis'mennost' narodov Severa: III: Jazyki i pi'smennost' paleoaziatskix narodov*. Moscow/Leningrad: Uchpedgiz. 181–222.
- Krejnovich E A (1937). *Fonetika nivxskogo (giljatskogo) jazyka*. Moscow/Leningrad: Uchpedgiz.
- Krejnovich E A (1955). 'Giljatsko-tunguso-man'chzhurskie jazykovye paralleli.' In *Doklady i soobshcheniia instituta jazykoznanija VIII*. Moscow: Akademija Nauk SSSR. 135–167.
- Krejnovich E A (1979). 'Nivxskij jazyk.' In *Jazyki Azii i Afriki III: Jazyku drevnej perednej Azii (nesemitskie), Iberijsko-kavkazskie jazyki. Paleoaziatskie jazyki*. Moscow: Nauka. 295–329.
- Mattissen J (2003). *Dependent-head synthesis in Nivkh. A contribution to a typology of polysynthesis*. Amsterdam: John Benjamins.
- Nedjalkov V P, Otaina G A & Xolodovich A A (1969). 'Morfologicheskij i leksicheskij kauzativ v nivxskom jazyke.' In Xolodovich A A (ed.) *Tipologija kauzativnyx konstruksij. Morfologicheskij kauzativ*. Leningrad: Nauka. 179–199.
- Nedjalkov V P & Otaina G A (1983). 'Rezul'tativ i kontinuativ v nivxskom jazyke.' In Nedjalkov V P (ed.) *Tipologija rezul'tativnyx konstruksij (rezul'tativ, stativ, passiv, perfekt)*. Leningrad: Nauka. 80–89.
- Otaina G A (1978). *Kachestvennye glagoly v nivxskom jazyke*. Moscow: Nauka.
- Panfilov V Z (1962). *Grammatika nivkhskogo jazyka Chast' I*. Moscow/Leningrad: Akademija Nauk SSSR.
- Panfilov V Z (1965). *Grammatika nivkhskogo jazyka Chast' II*. Moscow/Leningrad: Akademija Nauk SSSR.
- Panfilov V Z (1968). 'Nivxskij jazyk.' In *Jazyki Narodov SSSR V: Mongol'skie, tunguso-man'čžurskie i paleoaziatskie jazyki*. Moscow: Akademija Nauk SSSR. 408–434.
- Savel'jeva V N & Taksami C M (1970). *Nivxsko-russkij slovar'*. Moscow: Sovetskaja Entsiklopedija.
- Shternberg L J (1900). 'Obrazcy materialov po izucheniju giljatskogo jazyka i fol'klora, sobrannyx na ostrove Saxaline i v nizov'jax Amura.' *Izvestija imperatorskoj Akademii nauk. V serija* (tom XIII (4)). St. Petersburg: Akademija Nauk. 386–434.
- Tailleux O G (1960). 'La place du ghiliak parmi les langues paléosibériennes.' *Lingua IX*, 113–147.

Norse and Icelandic

T Eythórsson, University of Iceland, Reykjavík, Iceland

© 2006 Elsevier Ltd. All rights reserved.

Terminology and Historical Relations

Norse, or more specifically Old Norse, is a branch of medieval North Germanic (see **Germanic Languages; Indo-European Languages**). Old Norse is another name for Old West Nordic, referring to the language spoken from about 800 A.D. to the late 14th century in Norway and to the mid-16th century in Iceland. It was also spoken in the Faroe Islands, and in the Norse (Viking) settlements in the British Isles and Greenland. In a narrower sense, Old Norse is often used interchangeably with Old Icelandic, since most of the transmitted texts were written in Iceland. Modern Icelandic is the language of Iceland from the mid-16th century onwards, and is spoken today by about 290 000 people.

The earliest documented stage of North Germanic is Ancient Nordic (also known as Proto-Nordic, Runic

Norse, or Early Runic), attested in runic inscriptions in the Old Germanic writing system, the *futhark*, dating from about 150 to 800 A.D.. Towards the end of this period North Germanic began to divide into Old Norse (gradually splitting into Old Norwegian and Old Icelandic), on the one hand, and Old East Nordic (Old Danish, Old Swedish, and Old Gutnish), on the other.

Despite being in origin a West Nordic language like Icelandic and Faroese, Modern Norwegian has developed characteristics that are closer to Danish and Swedish (see **Norwegian**). Therefore, the modern Nordic (or Scandinavian) languages can be grouped into Mainland Scandinavian (Swedish, Danish, and Norwegian) and Insular Scandinavian (Icelandic and Faroese).

From Old Norse to Modern Icelandic

The evidence for Old Norse almost exclusively comes from Norway and Iceland; early documentation of Faroese is much poorer, and that of Norse in other

areas settled by Norsemen is very scanty. Early on, linguistic innovations started to separate Old Norwegian from Old Icelandic, but the differences between them did not become significant until the 14th century. The oldest Icelandic manuscripts do not display dialect variation, which may indicate leveling of putative preexisting dialect differences among the settlers, most of whom are reported to have come from Norway, either directly or via the British Isles. The most copious source of evidence involves prose texts transmitted in manuscripts in the Latin script, the earliest dating from the 12th century. Particularly important for the study of Old Icelandic phonology is the so-called First Grammatical Treatise (ca. 1150), an outstanding work in terms of its scientific precision and methodological rigor. A further source is found in two types of Old Norse poetry, the 'eddic' and 'scaldic' poems. Both types of poetry preserve numerous archaic features, in part due to their metrical form, and thus they represent a linguistic stage predating the earliest written texts. Finally, there are runic inscriptions in Old Norse, mostly from Norway but also from Iceland. There has been an unbroken written tradition from Old Norse to Modern Icelandic.

Phonology

Ancient Nordic had five vowels, which could be long or short, and three diphthongs. Due to umlaut, breaking and other sound changes, however, Old Norse had, by the mid-12th century, developed a system of twenty-six vowel phonemes and three diphthongs. In this system not only vowel quantity but also nasality were distinguished, but the latter distinction (which only pertained to long vowels) was lost early on. In the following centuries some further changes occurred in the vowel system, the most important being the loss of distinctive vowel length. This change, known as the 'great quantity shift,' took place in Icelandic in the 16th century, but in the three preceding centuries it had affected most of the other Nordic languages. As a result of these changes, the Modern Icelandic vowel system, consisting of eight vowel phonemes and five diphthongs, is very different from that of Old Norse, although the effects of the changes are obscured by conservative orthography. By contrast, the consonant system has remained more stable from Old Norse to Modern Icelandic. Certain characteristics of Modern Icelandic consonants are rare in other languages, e.g., preaspiration of voiceless stops (*happ* [hahp] 'luck') and devoicing of sonorants (*milt* [milt] 'mild,' *fínt* [fiɲt] 'fine'). However, such sounds also occur in Faroese and some Norwegian dialects, which may point to a common origin in Old Norse. Moreover, similar characteristics

in neighboring languages, Scots Gaelic and North Sámi (Saami, Northern), are possibly due to contact between these languages and Old Norse.

Morphology

Old Norse is distinguished from its Germanic relatives by two notable morphological innovations. The first involves the development of a definite article, which can be free or suffixed to the noun; in either case the noun and the article both inflect. There is no indefinite article in Old Norse, a situation which has been preserved in Modern Icelandic but not in the other modern Nordic languages. The second innovation is 'middle' verbs characterized by an ending which originated as an enclitic reflexive pronoun. The middle has various meanings, such as reflexive, reciprocal, anticausative, and passive. The last meaning is uncommon in Old Norse and Modern Icelandic, but has become the dominant one in Mainland Scandinavian.

Otherwise, most morphological categories known from other Old Germanic languages recur in Old Norse and, by and large, in Modern Icelandic as well. These include the four cases of nouns, pronouns, and adjectives, and three degrees of comparison for most adjectives and some adverbs. Nouns are inherently masculine, feminine, or neuter, but pronouns and adjectives agree in gender with the noun they modify. Finite verbs are inflected in three persons, two tenses (present, past), and three moods (indicative, subjunctive, imperative). The nonfinite verb forms comprise the infinitive and the present and past participles.

There are two numbers (singular and plural) in Old Norse, with the sole exception of the first and second person personal and possessive pronouns that preserve the dual number as well. In Icelandic the pronominal dual replaced the plural forms, whereas the old plural forms were restricted to honorific (formal) usage; in the past few decades the use of the honorific forms has decreased so that they are by now obsolete.

Syntax

Old Norse is a verb second (V2) language: the finite verb obligatorily occurs no later than in second position in all main clauses. V2 is an innovation in Germanic *vis-à-vis* other Indo-European languages, which may originally have been limited to certain main clause types and then generalized to all main clauses. Moreover, in Old Norse the finite verb occurs in second position in subordinate clauses as well, presumably due to an extension of the main-clause pattern.

This pattern ('symmetric V2') has been preserved in Modern Icelandic but not in Mainland Scandinavian.

Old Norse also has verb-initial clauses (V1), e.g., in direct questions, commands (imperatives), and conditional clauses. Declarative main clauses exhibiting V1 are frequent in narrative contexts ('narrative inversion').

In noun phrases (NPs), adjectival and pronominal modifiers regularly precede the head noun in Old Norse, but they may also follow it. Possessive genitives, on the other hand, generally follow the head noun, and this is also the case in Modern Icelandic.

In Old Norse the order of the nonfinite verb relative to an object in the verb phrase (VP) can be either verb-object (VO) or object-verb (OV). Such variable VP order also occurs in Old English. In Modern Icelandic OV orders were lost rather abruptly in the beginning of the 19th century, several hundred years later than in most other Nordic languages. The only exceptions to the strict VO pattern in Modern Icelandic are found with negative objects, which obligatorily precede the verb, and quantified objects, where either order may occur.

In the neutral word order in both main and subordinate clauses the subject occurs initially, immediately followed by the finite verb in second position (V2). In clauses containing a fronted nonsubject, the subject regularly follows the finite verb, although it may also occur further to the right, and even be extraposed to the end of the clause. Topicalization of VPs is ungrammatical in Modern Icelandic and seems not to be attested in Old Norse prose. However, fronting of nonfinite verbs (past participles, infinitives), as well as other head-like elements, is very common. This is 'stylistic fronting,' which only occurs in clauses that do not contain an overt subject ('subject gap').

As in other Germanic languages with morphological case, in Old Norse and Modern Icelandic subjects are typically in nominative case, direct objects in accusative case, and indirect objects in dative case. Various other patterns exist, however; in particular, dative objects are more common than in other Germanic languages. A further noteworthy characteristic is the occurrence of oblique (or 'quirky') subjects in Modern Icelandic (and Faroese). The status of the corresponding oblique NPs in Old Norse has been a matter of some controversy, but it is clear that in some respects they pattern syntactically with nominative subjects rather than with unambiguous objects. There is a tendency among some speakers of Icelandic to generalize dative case at the expense of accusative on some oblique subjects ('dative sickness'), or to replace the oblique case by nominative ('nominative substitution').

The emergence of a nonreferential or 'expletive' element, homonymous with the singular neuter pronoun *það* 'it,' was an innovation that gained ground in Icelandic in the latter half of the 18th century. About the same time, referential null arguments (*pro*-drop), occurring under certain rather well-defined conditions in Old Norse and early Modern Icelandic, were largely lost. A further innovation, apparently much more recent, is the 'new impersonal construction' (also called the 'new passive'), which has passive verb morphology (an auxiliary verb 'be' plus a past participle) but an accusative object in post-verbal position. This phenomenon does not seem to have a match in other Germanic languages, but there are typological parallels further afield, e.g., in Polish and Irish (Gaelic, Irish).

In archaic Old Norse (eddic and scaldic poetry), negation was expressed by a verbal prefix *ne* 'not,' also found in other Old Germanic languages. At this early stage of Old Norse the prefix could cooccur with a suffix (-*a* or -*at*) attaching to the finite verb, so that together the prefix and the suffix formed a discontinuous negation. These forms were lost early on, and negation came to be expressed by sentence adverbs (cf. Icelandic *ekki*, Danish *ikke* 'not').

Vocabulary

The Old Norse vocabulary belongs to the common Germanic stock. In addition to the inherited material, it contains a number of cultural loans from neighboring languages that the Vikings were in contact with. In Icelandic a few Celtic loanwords date from the earliest period, which is comprehensible in light of the fact that some of the Norse settlers came to Iceland via the British Isles, where they had been living in the proximity of Celts. Christianity, introduced in Iceland in the year 1000, led to an influx of new loanwords relating to religious and scholarly concepts, either directly from Latin or through intermediaries, especially Old English and Old Saxon. Further contact, especially trade and translations of the literature of chivalry, brought new loans from Low German that were mostly transmitted via Norwegian, although some of them were ultimately of French or Latin origin. In the wake of the Reformation in the mid-16th century more loanwords entered the language, in particular from High and Low German. The transmission mostly followed via Danish, which remained the most influential foreign language in Iceland until the middle of the 20th century. Since then the influence of English has been increasing steadily. A recent survey indicates that today, English is used more on a daily basis in Iceland than in any other Nordic country.

Despite long-standing contact with other languages, the vocabulary of Modern Icelandic is closer to Old Norse than that of any other Nordic language. This situation has in part been achieved by conscious effort. There is a long tradition of countermovement, dating back at least to the 18th century, against the infiltration of the language by foreign words and expressions, and language purism has been practiced as an active language policy since the mid-19th century. According to this policy, neologisms are created for new concepts (e.g., *sími* 'telephone,' *tölva* 'computer') rather than adopting words from other languages. Moreover, attempts have been made to resist phonological, morphological, and syntactic changes, even by resurrecting some archaic patterns; while some of these attempts have been quite successful, others have failed. The old custom of patronymics (*Höskuldsson* 'son of Höskuldur,' *þórhallsdóttir* 'daughter of þórhallur') has been preserved to a large degree, with the active support of the public authorities. The use of metronymics is also an option, although much less common, but it seems to have increased somewhat in recent years (*Guðrúnardóttir* 'daughter of Guðrún,' *Hrafnbjargarson* 'son of Hrafnbjörg'). In the Icelandic telephone directory people are generally listed under their first name.

Archaism and Innovation in Icelandic

Icelandic is often claimed to be a 'conservative' language that has preserved many archaic features, and it is probably true that it is no more difficult for the modern Icelander to read the Old Icelandic 13th-century sagas than it is, for example, for speakers of English to read Shakespeare. Moreover, despite a certain amount of geographically distributed language variation, which mainly affects pronunciation, there appear never to have been well-defined local dialects in Iceland with numerous distinctive characteristics. Nor are there strong contrasts between a standard and substandard register, at least compared to many other countries. Some possible reasons for this stability may be the geographic isolation of the country, an active conservative literary tradition, and a strong tradition of language purism. Nevertheless, as indicated above, numerous innovations have taken place since Old Norse, mostly affecting the phonology and syntax of the language rather than its morphology.

Bibliography

Árnason K (1980). *Quantity in historical phonology: Icelandic and related cases*. Cambridge: Cambridge University Press.

- Bandle O *et al.* (2002). *The Nordic languages: an international handbook of the history of the North Germanic languages*, vol. 1. Berlin: de Gruyter.
- Barðdal J (2001). *Case in Icelandic: a synchronic, diachronic, and comparative approach*. (Doctoral dissertation.) Lund: Institutionen för nordiska språk.
- Barðdal J & Eythórsson T (2003). 'The change that never happened: the story of oblique subjects.' *Journal of Linguistics* 39, 439–472.
- Benediktsson H (1972). *The first grammatical treatise*. Reykjavík: Institute of Nordic Linguistics.
- Benediktsson H (2002). *Linguistic studies, historical and comparative*. Thórhallsdóttir G, Thráinsson H, Friðjónsson J G & Ottósson K (eds.) Reykjavík: Institute of Linguistics.
- Einarsson S (1945). *Icelandic: grammar, texts, glossary*. Baltimore, MD: Johns Hopkins University Press.
- Eythórsson T (1995). *Verbal syntax in the Early Germanic languages*. Doctoral dissertation, Cornell University.
- Eythórsson T (2002). 'Changes in subject case marking in Icelandic.' In Lightfoot D (ed.) *Syntactic effects of morphological change*. Oxford: Oxford University Press. 196–212.
- Eythórsson T (2002). 'Negation in C: the syntax of negated verbs in Old Norse.' *Nordic Journal of Linguistics* 25, 190–224.
- Faarlund J T (1990). *Syntactic change: toward a theory of historical syntax*. Berlin: Mouton de Gruyter.
- Faarlund J T (2004). *The syntax of Old Norse*. Oxford: Oxford University Press.
- Guðmundsson H (1972). *The pronominal dual in Icelandic*. Reykjavík: Institute of Nordic Linguistics.
- Hansson G Ó (1999). 'Remains of a submerged continent: preaspiration in the languages of northwest Europe.' In Brinton L J (ed.) *Historical linguistics 1999*. Amsterdam: John Benjamins. 157–173.
- Helgason P (2002). *Preaspiration in the Nordic languages: synchronic and diachronic aspects*. Doctoral dissertation, Stockholm University.
- Hróarsdóttir T (2000). *Word order change in Icelandic: from OV to VO*. Amsterdam: John Benjamins.
- Jónsson J G (1996). *Clausal architecture and case in Icelandic*. Doctoral dissertation, Amherst: University of Massachusetts.
- König E & van der Auwera J (eds.) (1944). *The Germanic languages*. London: Routledge.
- Maling J & Sigurjónsdóttir S (2002). 'The "new impersonal" construction in Icelandic.' *Journal of Comparative Germanic Linguistics* 5, 97–142.
- Maling J & Zaenen A (eds.) (1990). *Syntax and Semantics 24: Modern Icelandic syntax*. San Diego: Academic Press.
- Ottósson K (1992). *The Icelandic middle voice: the morphological and phonological development*. Doctoral dissertation, University of Lund.
- Rögnvaldsson E (1995). 'Old Icelandic: a non-configurational language?' *NOWELE* 1995, 3–29.
- Rögnvaldsson E (1996). 'Word order variation in the VP in Old Icelandic.' *Working Papers in Scandinavian Syntax* 58, 55–86.

- Sigurðsson H Á (1989). *Verbal syntax and case in Icelandic*. Doctoral dissertation, University of Lund.
- Sigurðsson H Á (1993). 'Argument drop in Old Icelandic.' *Lingua* 89, 247–280.
- Thráinsson H (1979). *On complementation in Icelandic*. New York: Garland.

- Thráinsson H (1994). 'Icelandic.' In König & van der Auwera (eds.) 142–189.
- Thráinsson H (1996). 'Is Icelandic a natural language?' In Ottósson K G, Fjeld R V & Torp A (eds.) *The Nordic languages and modern linguistics* 9, 9–36.

North Philippine Languages

L A Reid, University of Hawai'i, Honolulu, HI, USA

© 2006 Elsevier Ltd. All rights reserved.

There are reported to be around 70 languages spoken in the north of the Philippines, nearly half of the total number of Philippine languages (Grimes and Grimes, 2000). For the purpose of this article, these languages include all of those spoken to the north of Manila, on the island of Luzon, and on the islands of the Batanes group, located in the Bashi Channel between Luzon and Taiwan. All of these languages belong to the branch of the Austronesian family, commonly referred to as Malayo-Polynesian (but in some works as Extra-Formosan), that began with the movement of Austronesian Neolithic seafaring people from what is now called Taiwan to eventually settle all of the Philippines and ultimately the rest of the Pacific. Recent archaeological evidence (Bellwood *et al.*, 2003: 158) suggests that this movement began around 4000 years ago, with a small group of Austronesian people leaving the eastern coast of Taiwan, settling the Batanes Islands and eventually reaching the northern coast of Luzon. The Philippines was already occupied at that time by a large number of Negrito bands of hunter-gatherers, most of which are already extinct or which have been completely assimilated by the technologically superior immigrating Austronesians. There are still, however, more than 20 groups of Negritos located in relatively remote areas of the northern Philippines. About 15 of these groups, variously called Agta, Alta, and Arta, live in and around the Sierra Madre mountain range, with five or more groups called Ayta scattered around the Sambal mountain range in the west of Luzon. All Negrito groups adopted the language of their closest Austronesian neighbors in the distant past (Reid, 1987), but today they have diverged to the point where they speak languages which are clearly distinct from those of their neighbors. The Negrito languages are now highly endangered, because of the continuing pressure to integrate with their non-Negrito neighbors and the influence of the local trade languages.

Three of the languages in the northern Philippines, Ilokano (Ilocano) (with 6 636 000 speakers, or approximately 8.7% of the total Philippine population of 76 504 000, according to the National Statistics Office of the Philippines 2000 Census of Population and Housing), Kapampangan (Pampangan) (with 2 066 800 speakers, or 2.7%), and Pangasinan (with 1 185 600 speakers, or 1.5%) are referred to as 'major' languages, based on number of speakers. Of these, Ilokano is most important, being widely spoken as a second language and used as a language of wider communication throughout the north of Luzon. These three languages belong to three different sub-groups of Philippine languages. Ilokano is a first-order branch of the major language family in the north of the Philippines, referred to as northern Luzon or Cordilleran (Reid, 1979). Pangasinan is a member of the Central–Southern branch of the same family, while Kapampangan groups with the Sambalic languages, and possibly ultimately with the Batanic languages Itbayat and Ivatan, spoken in the far north of the Philippines (and their closely related language, Yami, spoken on Lányü or Orchid Island in Taiwan).

Tagalog has been used for a long period as a lingua franca in some areas of the northern Philippines, particularly on the eastern coast of Luzon, as far north as Paranan, with extensive influence on languages spoken in the area. In addition, in the guise of Filipino, the national language of the Philippines, the language has in recent years been moving out of the classroom throughout northern Luzon and into the daily lives of the younger generation, competing with Ilokano as a tool for communicating with outsiders.

The northern languages of the Philippines have a type of syntax that is found throughout the country (Reid and Liao, 2004). Predicates typically occur at the beginning of constructions, only allowing topicalized NPs to precede them. Noun phrases are typically introduced by one of a series of short, often monosyllabic forms which specify semantic features of the lexical head of the phrase, whether it is a common or personal noun, whether it is singular or plural, and in some languages its spatial or temporal relationship to

the speaker. This form may also mark the case of the noun. Nominative (or absolutive) NPs are typically unmarked for case when they are lexical nouns, but are case marked when a pronominal substitute replaces them.

Each of the languages is morphologically ergative, with the patientlike noun phrase of transitive sentences being marked in the same way as the single argument of intransitive sentences, while the agentlike noun phrase of transitive sentences is marked in the same way as that of genitive constructions within a noun phrase.

Intransitive constructions are either monadic (having a single core argument) or dyadic (having two core arguments). The latter construction is amenable to an antipassive analysis, with the patientlike oblique argument being typically indefinite and marked with the same case marking as that used for locative noun phrases, unlike Tagalog, which marks such oblique NPs as genitives. Intransitive constructions are often referred to in the literature as ‘actor-focus,’ in that the morphology on the verb conveys the information that the nominatively marked NP is an ‘actor.’ Transitive constructions are typically considered to constitute at least four different types, ‘goal focus,’ ‘locative focus,’ ‘instrument focus,’ and ‘benefactive focus,’ depending on the morphology of the verb. Structurally these constructions are identical. They differ only in the semantic interpretation of the nominative NP of the sentence.

The northern languages of the Philippines tend to be more complex phonologically than other Philippine languages, which typically have only four or five vowels. Casiguran Dumagat Agta, for example, has developed an eight-vowel system, while Balangaw (Balangao), a Central Cordilleran language, is reported to have seven vowels. Karao, a southern Cordilleran language, has developed a series of fricatives /f/, /θ/, and /x/, which are unusual in Philippine languages. Several of the languages have very complex morphophonemic (sandhi) alternations. These include Ibanag, a northern Cordilleran language (Brandes and Scheerer, 1927–1928), and Karao and its sister language, Inibaloi (Ibaloi) (Brainard, 1994). Some dialects of Bontok (Bontoc), Ifugao, and Kalinga exhibit a complex range of often voiceless fricative or affricate prevocalic variants of their voiced stops, /b/, /d/, and /g/, such as Guinaang Bontok [f], [ts], and [k^h] (Himes, 1984–1985), which because of the influence of English in the schools are losing their

environmental conditioning and are now becoming separate phonemes in the languages (Reid, 2005).

Published text resources are available on a number of Cordilleran languages. Moreover, one of the finest dictionaries of a Philippine language is that of the central Cordilleran language Ifugao (Newell, 1993).

Bibliography

- Barlaan R R (1999). *Aspects of focus in Isnag*. Manila: Linguistic Society of the Philippines.
- Bellwood P, Anderson A, Stevenson J & Dizon E (2003). ‘Archaeological and palaeoenvironmental research in Batanes and Ilocos Norte provinces, northern Philippines.’ *Bulletin of the Indo-Pacific Prehistory Association* 23, 141–161.
- Brainard S (1994). *The phonology of Karao, the Philippines*. Canberra: Pacific Linguistics.
- Brandes C O & Scheerer O (1927–1928). ‘On sandhi in the Ibanag language.’ In Scheerer O (ed.) *The Archives* 6. Manila: University of the Philippines. 5–42.
- Grimes BF & Grimes JE (eds.) (2000). *Ethnologue: languages of the world* (14th edn.). Dallas: SIL International.
- Himes R S (1984–1985). ‘Allophonic variation and the Bontok-Kankanaey voiced stops.’ In *Philippine Journal of Linguistics*. 15–16, 49–56.
- Himes R S (1998). ‘The southern Cordilleran group of Philippine languages.’ *Oceanic Linguistics* 37(1), 120–177.
- Newell L E (1993). *Batad Ifugao dictionary with ethnographic notes*. Manila: Linguistic Society of the Philippines.
- Reid L A (1974). ‘The central Cordilleran subgroup of Philippine languages.’ *Oceanic Linguistics* 13, 511–560.
- Reid L A (1979). ‘Towards a reconstruction of the pronominal systems of Proto-Cordilleran, Philippines.’ In Nguyen D L (ed.) *Southeast Asian linguistic studies* 3. Canberra: Research School of Pacific Studies, Department of Linguistics, Australian National University. 259–275.
- Reid L A (1987). ‘The early switch hypothesis: linguistic evidence for contact between Negritos and Austronesians.’ *Man and Culture in Oceania* 3(Special Issue), 41–60.
- Reid L A (2005). ‘A cross-generational view of contact-related phenomena in a Philippine language: phonology.’ In Quakenbush S & Dayag D (eds.) *Sociolinguistics and language education in the Philippines and beyond: festschrift in honor of Ma. Lourdes S. Bautista*. Manila: Linguistic Society of the Philippines/Summer Institute of Linguistics.
- Reid L A & Liao H-C (2004). ‘A brief syntactic typology of Philippine languages.’ *Language and Linguistics* 5(2), 433–490.

Norwegian

T A Åfarli, Norwegian University of Science and Technology, Trondheim, Norway

© 2006 Elsevier Ltd. All rights reserved.

Sociohistorical Setting

Norwegian together with Danish and Swedish constitute the Mainland Scandinavian languages, which, together with the Insular Scandinavian languages, Faroese and Icelandic, constitute the Scandinavian languages. The Scandinavian languages belong to the Germanic family of the Indo-European languages.

Norwegian is exceptional in having two officially recognized written standards. These are called *Bokmål* 'book language' and *Nynorsk* 'New Norwegian'. Bokmål is used by more than 80% of the population, whereas Nynorsk is used by less than 20%, mainly in the area stretching from the interior of southern Norway to the western coast. As for the spoken language, there is a rich variety of dialects, although all of them are mutually intelligible. The use of (nonstandard) dialects is widely accepted, even in more formal contexts. There is no officially recognized standard for the spoken language. It is important to note that Bokmål and Nynorsk are written standards that individuals choose largely irrespective of which dialectal variety they speak.

The two written standards have their background in the history of Norway. From 1380 to 1814, Norway was in a political union with Denmark, and Danish was used as a written language. Eventually, during the flourishing of nationalism after 1814, a Dano-Norwegian written standard was developed in the 19th century, bringing Norwegian elements into the Danish language. This standard gradually evolved into present-day Bokmål. Also, in the 19th century, a more radical approach was followed by Ivar Aasen (1813–1896). He developed a written standard based on the spoken dialects. This standard gradually evolved into present-day Nynorsk.

Today, Bokmål and Nynorsk are quite similar, but there are still certain spelling differences both regarding content words and grammatical morphology. In the description of Norwegian below, I will concentrate on the common structural features of the spoken language as a whole. All examples are written in the Nynorsk standard unless otherwise stated.

Morphology and Phonology

Finite verbs in Norwegian only show tense distinctions (past, present). There are two nonfinite verb forms, the infinitive and the past participle, and an adjectival present participle form. Example (1)

shows the inflectional paradigm for the weak verb *kjøpe* 'buy'.

- (1) *kjøpe* (infinitive) - *kjøper* (present) - *kjøpte* (past)
- *kjøpt* (past participle)

Nouns have no productive case distinctions. However, they inflect for number (singular, plural) and definiteness (definite, indefinite), and there are three genders (masculine, feminine, neuter). Example (2) shows the paradigm for the feminine noun *geit* 'goat'.

- | | | |
|-------|-------------------------|----------------------------|
| (2) | sg | pl |
| indef | <i>geit</i> 'goat' | <i>geiter</i> 'goats' |
| def | <i>geita</i> 'the goat' | <i>geitene</i> 'the goats' |

As shown, definiteness is marked inflectionally as a suffix (*-ene* is a portmanteau morph expressing both definiteness and plurality). There is also a free-form indefinite article, varying according to gender; *ei* is the feminine form, as in *ei geit* 'a goat'.

Adjectives show agreement in gender (m/f and n) and number (sg, pl) in predicative and attributive positions.

As for phonological properties, Norwegian dialects have relatively rich vowel and consonant systems. Also, Norwegian dialects make extensive use of diphthongs. In most dialects, tones may have distinctive function.

Syntax

Norwegian is a SVO language with fixed word order, cf. (3).

- (3) *Mannen kjøper geita.*
'The man buys the goat.'

Norwegian is also a verb second (V2) language, with the finite verb in second position in main declarative clauses. Thus, topicalization of the direct object in (1) yields *geita kjøper mannen* 'the goat the man buys' with the finite verb in the V2 position.

Auxiliary verbs are positioned between the subject and the main verb, as in *mannen har kjøpt geita* 'the man has bought the goat'. Finite auxiliary verbs follow the V2 pattern, cf. *geita har mannen kjøpt* 'the goat the man has bought'.

Adverbials are positioned in the middle field (sentence adverbials) or toward the end of the clause (predicate adverbials).

- (4) *Mannen har aldri kjøpt geiter på ein søndag.*
'The man has never bought goats on a Sunday.'

Like other V2 languages, Norwegian shows an asymmetry between main and embedded clauses as to the

relative distribution of sentence adverbial and finite verb, compare (4) to the embedded clause in *eg veit at mannen aldri har kjøpt geiter på ein søndag* 'I know that the man has never bought goats on a Sunday'.

Norwegian has a strict subject requirement in finite clauses. If there is no semantic subject, an expletive subject must be inserted.

(5a) Det regnar.
'It rains.'

(5b) Det står ei flaske på bordet.
it stands a bottle on table-the
'There is a bottle standing on the table.'

In most dialects, *det* 'it' is used as expletive subject in both meteorological (5a) and presentational (5b) sentences. Some dialects allow *der* 'there' as an expletive subject in addition to *det* 'it'.

Norwegian has two main types of passive, namely periphrastic passive and reflexive (*s-*) passive. Example (6b) is written in the Bokmål standard.

(6a) Geita blir kjøpt (av mannen) i dag.
goat-the becomes bought (by man-the) to day
'The goat is bought by the man today.'

(6b) Geita kjøpes (av mannen) i dag.
goat-the buy-s (by man-the) to day
'The goat is bought by the man today.'

Both types allow impersonal passive, as e.g., in *det blir kjøpt ei geit i dag* 'there is bought a goat today'.

There are two main types of interrogative clauses: yes/no questions (formed by placing the finite verb in initial position) and questions with a question word (with the question word placed in initial position). As for relative clauses, the most common type is introduced by the complementizer *som*. The complementizer is optional if the relativized position is a nonsubject, as in *geita (som) mannen kjøpte* 'the goat (that) the man bought'. Infinitival clauses are introduced by the infinitive marker *å* 'to', as in *mannen prøver å kjøpe geita* 'the man tries to buy the goat'.

Bibliography

- Askedal J O (1994). 'Norwegian.' In König E & van der Auwera J (eds.) *The Germanic languages*. London: Routledge. 219–270.
Faarlund J T, Lie S & Vannebo K I (1997). *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Nostratic Hypothesis

A Dolgopolsky, University of Haifa, Haifa, Israel

© 1994 Elsevier Ltd. All rights reserved.

Nostratic is a hypothetical macro-family of languages, embracing Indo-European, Afro-Asiatic, Kartvelian, Uralic, Altaic, and Dravidian. The hypothesis is based on a large number of common roots (more than 2000 known in the early 1990s) and many common grammatical morphemes (pronouns and auxiliary words which later became prefixes and suffixes in the descendant languages), in which regular sound correspondences and results of regular phonological changes are observed. The common roots include basic lexical items, e.g., **ǵāso* 'stay, be' (in Indo-European [**es-* 'be'], Afro-Asiatic, Kartvelian, Uralic), **wete* 'water' (all branches except Kartvelian), **ǵitā* 'eat' (in Indo-European, Afro-Asiatic, Altaic), **bari* 'take' (all branches except Uralic), **2eyV* 'come' (in Indo-European [**ei-* 'go'], Afro-Asiatic, Uralic, Altaic, Dravidian), **nimru* 'name' (Indo-European, Afro-Asiatic, Uralic, Altaic), as well as words connected with cultural conditions of

the ancient (presumably final paleolithic) society, e.g., **kālu* 'woman of another exogamic moiety' (> 'sister- or daughter-in-law,' 'bride'; present in all branches, e.g., Indo-European **ǵlw-* 'sister or daughter-in-law' > Latin *glōs*, Greek *galōs*, Slavic **zolv-*), pronouns: **mi* 'I', preserved as a pronoun or as a morpheme of 1sg in almost all branches; **t'ü* > **ti* 'thou,' preserved in Indo-European, Afro-Asiatic, Uralic, and Altaic; **k'o* 'who' (in Indo-European, Uralic, and Altaic), **mi* 'what' (in Afro-Asiatic, Uralic, Altaic, Dravidian, and Kartvelian).

The parent language had, most probably, an analytic grammatical structure with a strict word order (sentence-final predicate; object preceding the verb; nonpronominal attribute preceding the head; a special position for unstressed pronouns) and with grammatical meanings expressed by word order and auxiliary words (e.g., postpositions: **nu* for genitive, **ma* for marked accusative, and others). In the descendant languages this analytic grammar evolved towards a synthetic one. The phonological system (reconstructed by V. Illich-Svitych (1971–84) and A. Dolgopolsky (1989) in the framework of a

Nostratic historical phonology) included a rich consonantism (with threefold opposition of voiced/voiceless/glottalized [ejective] stops and affricates, with three series of sibilants and affricates, with lateral obstruents, laryngeal, pharyngeal, and uvular consonants), and a vowel system of 7 vowels. The ancient Nostratic parent language seems to have existed in the preneolithic period (up to ca. 15 000 or 12 000 BC) somewhere in southwest Asia. But most descendant proto-languages (e.g., Proto-Indo-European) existed during the neolithic period (with agriculture and

husbandry, resulting in a demographic explosion, which can explain their spread throughout Eurasia and the northern half of Africa).

Bibliography

- Dolgopolsky A (1989). Problems of Nostratic comparative phonology. In Shevoroshkin V (ed.) *Reconstructing Languages and Cultures*. Bochum: Brockmeyer.
 Illich-Svitych V (1971–84). *Essay of Comparison of the Nostratic Languages*, 3 vols. Moscow: Nauka.

Nuristani Languages

A L Grünberg, St Petersburg Institute of Linguistics, St Petersburg, Russia

© 1994 Elsevier Ltd. All rights reserved.

Nuristani, sometimes known as Kafiri or Kafir languages (NL), are a group of Indo-European languages close in many aspects to the Dardic languages of the Indo-Iranian branch; but in some points of historical phonology (an early loss of aspiration, the reflexes of \hat{u} , \hat{g} , \hat{gh} in the form of homorganic affricates c , z ; the preservation of s after u) the NL differ from the Indo-Aryan languages. It is supposed that the Nuristani languages separated from the Indo-Iranian group before it split into the Indo-Aryan and Iranian.

The limits of the area where the NL are spoken coincide with the borders of the historical province of Nuristan (former Kafiristan) situated in the high mountains on the southern slope of the Eastern Hindu Kush (Afghanistan). Nuristan was nearly completely isolated from the outer world until the very end of the nineteenth century and then again until the 1930s.

There are five Nuristani languages: Kati, Kamviri, Ashkun, Waygali, and Prasun.

Kati is divided into two dialects. The western one is spoken in the valleys of Ramgel and Kulem, the two sources of the Alingar river, which is a right-bank tributary of the Kabul river; it is also spoken in the valley of Kantiwa, the right source of the Pech river, in its turn a right-bank tributary of the Kunar river (named Chitral in its upper part). The eastern dialect is spoken in the upper part in the valley of the Katigal (Bashgul, Landaisin) river, a left tributary of the Kunar river. The Ashkun language is also divided into at least two dialects: the western one is located in small valleys on the left side of the Alingar river; the eastern is spoken in Wama, a large village in the

Pech valley. The Waygali language occupies the valley of the Waygal river, a left tributary of the Pech river, also in the valley of Tregam in the same Pech region. There are at least three dialects in Waygali.

Kamviri is the language of a large Kamdesh village and some small villages in the middle part of the Bashgul valley, at a lower altitude than the Kati-speaking area. Prasun is spoken in a very isolated valley of Prasun, the left source of the Pech river, which divides the Kati-speaking area into two parts.

A number of Nuristanis speaking Kati, Waygali, and Kamviri now live in Kabul.

The first four languages (Kati, Kamviri, Ashkun, and Waygali) are closely related to each other, while Prasun occupies a specific position not only within the Nuristani group but in a sense also among the Indo-European languages as a whole.

The total number of people speaking NL is not known exactly, but probably does not exceed tens of thousands. A relatively great amount of Nuristanis (especially Kati-, Waygali- and Kamviri-speaking people) are bilingual, speaking Pashto or Dari as a second language.

The phonetic systems of NL contain a series of retroflex consonants, \check{s} , \check{z} , \check{t} , \check{d} , \check{c} , \check{j} , \check{n} , \check{r} . The palatalization of consonants is typical of all NL except Prasun. In Kati nearly all consonants have phonologically opposed palatalized pairs.

The noun has two case forms, direct and indirect. As a rule there is no morphological indication of number and gender in the direct case, but the masculine and feminine singular forms, and the plural form for both genders, are different in the indirect case.

All NL have relatively complicated systems of modal and temporal forms, as well as a great amount of various nonpersonal forms (participles, gerunds, absolutives, etc.).

All NL (except Prasun) have an ergative structure of sentences with a transitive verb in any past tense formed from the past stem; the verb agrees in gender, number, and person with the object. There is no ergative construction in Prasun, but there is a difference in conjugation of transitive and intransitive verbs in the past tense.

A structural feature specific to all NL is a very peculiar and sophisticated system of spacial orientation, determining the location of an object or the direction of its movement. The horizontal/vertical axes, certain objects on the earth's surface (for example a river or a mountain pass) as well as the subject of speech serve in this system as coordinates. Thus in Kati more than fifteen series of such abstract means of spacial orientation exist. Each series contains a preverb, three adverbs, and an adjective; so there are at least seventy-five abstract ways to locate an action of an object in the space. Still more complicated is such a system in Prasun, where theoretically more than a hundred ways of spatial orientation can be used.

The vocabulary of NL included, until the late twentieth century, only a very small amount of loanwords, but now a process of penetration of loanwords from Pashto and Dari is in progress everywhere in Nuristan and especially among the Nuristanis living in Kabul.

Bibliography

- Edelmann D I (1983). *The Dardic and Nuristani Languages*. Moscow.
 Grünberg A L (1971). Opyt lingvisticheskoy karty Nuristana. *Strany i Narody Vostoka* X.
 Grünberg A L (1980). *Jazyki Vostochnogo Gindukusha*. Moscow: Jazyk Kati.
 Grünberg A L & Steblin-Kamensky I M (1974). Etnolingvisticheskaya charakteristika Vostochnogo Gindukusha. In: *Problemy karto-grafirovaniya v jazykoznanii i etnografii*. Leningrad.

Nuuchahnulth

J Stonham, University of Newcastle, Newcastle upon Tyne, UK

© 2006 Elsevier Ltd. All rights reserved.

Nuuchahnulth, also known as Nootka, Nutka, Aht, Takhaht, and *t'aat'aaqsapa*, and by the various dialect designations, is one of the earliest documented languages of Western Canada, with first contact going back to the 1770s. Early work focused mostly on the vocabulary, with word lists being documented by a number of early explorers. The earliest detailed grammatical investigations include those of Knipe (1868), Boas (1890), and Sapir (1924). The name Nuuchahnulth is an anglicized version of *nuučaahnuł* 'along the mountains' (*nučĩ* 'mountain(s)', *-a'ñuł* 'along': ʔ indicates variable length of the vowel). Geographically, the Nuuchahnulth people occupy the west coast of Vancouver Island, from Brookes Peninsula in the north to Bamfield in the south.

Nuuchahnulth has a relatively large consonant inventory as shown in Table 1. Conversely, the vowel system is quite simple, involving just the three vowels, /i, u, a/, and a length distinction. There are, in addition, two mid vowels that are only encountered in the long variety, /e/ and /o/, at least phonologically, and even then only under certain special circumstances, such as in foreign borrowings. Primary stress

is predictable and appears on one of the first two syllables of a word, depending on weight.

Important morphophonological processes include glottalization and lenition, which affect the preceding consonant and depend on the individual suffix triggering the effect (abbreviations: LOC, locative; DUP, duplication; SUFF, suffix, MOM momentaneous aspect, NOW contemporaneous, MC momentaneous causative, SUB subordinate, L lengthening, R reduplication. All data are from Sapir's unpublished fieldnotes, (Sapir, no date), unless otherwise indicated.

- (1a) *hiitaht'ixšiʔaλ*
 hita-ht -ʔix -šiλ -ʔaλ
 LOC-exit woods-go for[L]-MOM-NOW
 'They started out of the woods.'
- (1b) *wiškum*
 wišk -ʔum
 angry -on the rocks
 'angry on the rocks'
- (1c) *ʔaayimkiλqas*
 ʔaya -miik -iλ -qaʔs
 many -getter of -MC -1s.SUB
 'may I be a getter of many ...'
- (1d) *titinkum*
 DUP-ti -ñuk^ω -im
 SUF-wipe -at the hand [R] -thing
 'handwiper'

Table 1 Nuuchahnulth consonant inventory

| Labial | Alveolar | Lateral | Alveopalatal | Velar/labiovelar | Uvular/labiouvular | Pharyngeal | Glottal |
|--------|----------|----------------|--------------|------------------|--------------------|----------------|---------|
| p | t | | | k | q | | ʔ |
| p̥ | t̥ | | | k̥ | q̥ | ħ | |
| | | | | k ^ω | q ^ω | | |
| | | | | k̥ ^ω | q̥ ^ω | | |
| | s | ʃ | š | x | x | ħ | h |
| | | ʃ ^ω | | x ^ω | x ^ω | ħ ^ω | |
| | c | ç | č | | | | |
| | č | ç' | č' | | | | |
| m | n | | y | w | | | |
| m̥ | n̥ | | y̥ | w̥ | | | |

As can be seen from (1a) and (1b), a glottalizing suffix will trigger glottalization of the preceding consonant, whereas a nonglottalizing suffix ((1c) and (1d)) will not. Other common morphophonological processes include labialization, delabialization, vowel coalescence, and ʔ-epenthesis. One further morphophonological process involves a class of vowels labeled variable length; such vowels appear long in the first foot of the word, but are short after that. These are characteristically Wakashan in nature and constitute a special class of vowels.

Morphologically, Nuuchahnulth is extremely complex, with upward of 500 bound derivational morphemes, along with numerous inflectional paradigms. The creative genius of the language is demonstrated by the following words:

- (2) *cicihʔaqχmapt* ‘crab-apple tree’ (< *cih* ‘sour’ -ʔaqχ ‘inside’ -mapt ‘plant’)
čičiyupkuk ‘spaghetti’ (< *čiyup* ‘intestines’ -*kuk* ‘resemble’)
niikmatiičakč ‘guitar’ (< *nik* ‘scratch’ -*mat* ‘about’ -*iiču* ‘tool’)
‘at strings’ -čak^ω

Morphological processes include reduplication, infixation, and suffixation, although lexical compounding is absent. There is a class of lexical suffixes that requires reduplication of some part of the root as a concomitant of the attachment (3a), in addition to reduplication indicating plurality (3b), distributivity (3c), repetition (3d), and other aspectual categories. In fact, double reduplication may occur in contexts in which both derivational and inflectional triggers to reduplication cooccur (DISTRIB, distributivity; REP, repetition; PASS, passive)

- (3a) *kuukuhiŋqit*
DUP-kuh -inŋit
SUF-hollow -at the ribs [RL]
‘with a hole in the side’
(3b) *taataayi*
DUP-taayii
PL-older brother
‘older brothers, seniors’

- (3c) *nunupqimtayiiʔat*
DUP-nup -qimt -ayiʔ -ʔat
DISTRIB-one -classifier -give -PASS
‘He gave a dollar to each.’

- (3d) *tuuxtuux^ωa*
DUP-tux -(y)aʔ
REP-jump -DUR
‘jumping’

Predicates are typically marked for aspect and often for location (e.g., -*ʔi* ‘in the house,’ -*ʔas* ‘on the ground, in the village,’ and -*ʔis* ‘on the beach’), and bear one of a set of paradigmatic mood/person/number suffixes and other possible markers. Tense is optionally marked on predicates, as is plural for nouns. Nouns may be marked for number, diminutive, augmentative, former/future state, and possession.

Syntactically, Nuuchahnulth may be described as head-initial and head-marking. The most common word order is verb initial, either verb–subject–object or verb–object–subject, although arguments may be omitted. In keeping with its head-marking nature, possessors follow possessees (4a) and relative clauses follow their heads (4b):

- (4a) *hawitukʔi* q^ωayačiiktaqimʔ
hawit-uk =ʔiʔ q^ωayačiik -taqimʔ
chief-POSS =DEF wolf -tribe
‘the chief of the wolves’
(4b) *tuucsmeʔi* yaq^ωacʔitq ʔaŋa
tuucsma-ʔiʔ yaq^ω-ac -ʔiʔtq ʔaŋa
woman-DEF REL-belong to -3s.REL child
‘the woman whose child he was’

In the case of relative clauses, there are two types – headed (4b), involving either *yaq^ω* ‘who(m)’ or *q^ωi* ‘which,’ or headless (5), in which the specifier *ʔiʔ* is attached to the predicate of a relative clause:

- (5) *qahšiχʔi*
qah -šiχ =ʔiʔ
die -MOM =DEF
‘the dead (ones)’

A process reminiscent of noun incorporation exists, but with a rather broad range of implementation:

- (6a) ʔuutʔaapʔaʔ $[[\text{suu} \text{h} \text{a} \text{a}]_{\text{N}}]_{\text{NP}}$
 ʔu - itʔaap - 'aʔ $\text{suu} \text{h} \text{a} \text{a}$
 REF-bring...as -NOW spring
 gift silver salmon
 'He brought a gift of silver spring salmon.'
- (6b) $\text{suu} \text{h} \text{ʔiitʔa} \text{p} \text{at}$ $[[\text{t} \text{ʔi}]_{\text{N}}]_{\text{NP}}$
 $\text{suu} \text{h} \text{a} \text{a}$ - itʔaap - 'at
 silver spring salmon -bring...as gift -PASS
 'He brought a gift of silver spring salmon.'

Elements incorporated may include numerals (7a), adverbs (7b), adjectives (7c), or nouns (6b), always from the object argument. Subjects of unaccusative verbs may also be involved (7d):

- (7a) $\text{h} \text{a} \text{a} \text{y} \text{u} \text{m} \text{i} \text{i} \text{k} \text{u} \text{k}$ $\text{ʔu} \text{h} \text{ʔi} \text{ʃ}$ $\text{m} \text{u} \text{u}$ $\text{ʔi} \text{i} \text{h} \text{t} \text{u} \text{u} \text{p}$
 $\text{h} \text{a} \text{y} \text{u}$ - $\text{m} \text{i} \text{k}^{\text{w}}$ - uk $\text{ʔu} \text{h} \text{ʔi} \text{ʃ}$ $\text{m} \text{u} \text{u}$ $\text{ʔi} \text{i} \text{h} \text{t} \text{u} \text{u} \text{p}$
 ten -capture -POSS and four whale
 'He captured fourteen whales.'
- (7b) $\text{ʔi} \text{i} \text{h} \text{i} \text{i} \text{t}$ $\text{x} \text{u} \text{t}$ $\text{ʔ} \text{a} \text{p} \text{a} \text{c}$
 $\text{ʔi} \text{i} \text{h}^{\text{w}}$ - $\text{i} \text{i} \text{t}$ $\text{x} \text{u} \text{t}$ $\text{ʔ} \text{a} \text{p} \text{a} \text{c}$
 very -make good canoe
 'He made a really nice canoe.'
 (Rose, 1981)
- (7c) $\text{ʔi} \text{i} \text{h} \text{n} \text{a} \text{a} \text{k}$ $\text{ʔ} \text{a} \text{p} \text{a} \text{c}$
 $\text{ʔi} \text{i} \text{h}^{\text{w}}$ - $\text{n} \text{a} \text{k}^{\text{w}}$ $\text{ʔ} \text{a} \text{p} \text{a} \text{c}$
 big -have canoe
 'He has a big canoe.'
 (Rose, 1981)
- (7d) $\text{ʔ} \text{a} \text{y} \text{a} \text{s} \text{u} \text{u} \text{x}$ $\text{n} \text{a} \text{m} \text{int} \text{ʔ} \text{a} \text{th}$
 $\text{ʔ} \text{a} \text{y} \text{a}$ - $\text{s} \text{a} \text{w} \text{i} \text{x}$ $\text{n} \text{a} \text{m} \text{int}$ - $\text{'a} \text{th}$
 many -die Namint-tribe
 'Many Namint people died.'

Coordination may involve noun phrases, verb phrases, or intermediate units and may employ one of several conjunctions, including $\text{ʔi} \text{ʃ}$ and $\text{ʔu} \text{h} \text{ʔi} \text{ʃ}$, both meaning 'and.' Subordination is marked by special inflectional paradigms and may or may not involve subordinating conjunctions such as $\text{ʔ} \text{a} \text{n} \text{i}$ 'that,' $\text{ʔ} \text{u} \text{y} \text{i}$ 'when,' etc. There is a marker, -ʔat , which has been labeled variously passive, switch reference, or discourse marker. (The debate over the role of this morpheme continues, for which see Kim (2004) and references therein.) For further information on Nuuchahnulth, the reader is referred to the bibliography appended to this article, in particular the discussions in Davidson (2002), Jacobsen (1979a,b), Kim (2003), Nakayama (2001), Rose (1981), and Stonham (1999, 2004).

Bibliography

- Boas F (1890). 'The Nootka.' In *Report of the 60th meeting of the British Association for the Advancement of Science* (BAAS-R 60). 582–604.
- Davidson M (2002). *Studies in Southern Wakashan (Nootkan) grammar*. Ph.D. diss., State University of New York at Buffalo.
- Jacobsen W H Jr (1979a). 'Wakashan comparative studies.' In Campbell L & Mithun M (eds.) *The Languages of Native America*. Austin, TX: University of Texas Press. 766–791.
- Jacobsen W H Jr (1979b). 'Noun and verb in Nootkan.' In Efrat B S (ed.) *The Victoria conference on northwestern languages*. Victoria, BC: British Columbia Provincial Museum. 83–155.
- Kim E-S (2003). *Theoretical issues in Nuu-chah-nulth phonology and morphology*. Ph.D. diss., University of British Columbia.
- Kim E-S (2004). 'The morphological status of -ʔat in Nuu-chah-nulth.' In *Proceedings of BLS 30*. Berkeley: University of California Press.
- Knipe C (1868). *Some Account of the Tahkaht Language, as Spoken by Several Tribes on the Western Coast of Vancouver Island*. London: Hatchard.
- Nakayama T (2001). *Nuuchahnulth (Nootka) morphosyntax*. Berkeley: University of California Press.
- Rose S M (1981). *Kyuquot grammar*. Ph.D. diss., University of Victoria.
- Sapir E (no date). *Fieldnotes on Nootka*. Boas collection of the American Philosophical Society [W2a.18]. Philadelphia, PA.
- Sapir E (1924). 'The rival whalers, a Nitinat story (Nootka text with translation and grammatical analysis).' *International Journal of American Linguistics* 3, 76–102.
- Sapir E & Swadesh M (1939). *Nootka texts, tales and ethnological narratives, with grammatical notes and lexical materials*. Philadelphia & Baltimore, MD: Linguistic Society of America.
- Sapir E & Swadesh M (1955). *Native accounts of Nootka ethnography*. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore and Linguistics.
- Stonham J (1999). *Tsishaath Nootka phonetics and phonology*. [Lincom Studies in Native American Linguistics 32.] Munich: Lincom Europa.
- Stonham J (2004). *Linguistic theory and complex words: Nuuchahnulth word formation*. London: Palgrave Macmillan.
- Swadesh M (1939). 'Nootka internal syntax.' *International Journal of American Linguistics* 9, 77–102.

Nyanja

S Mchombo, University of California, Berkeley, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

History and Politics

ChiNyanja (Nyanja) is a language of the Bantu group of the Niger-Kordofanian language family, and is spoken in parts of eastern, central, and southern Africa. It is spoken in Malawi, where, from 1968 until recently, under the name of Chichewa, it served as the national language. It is also spoken in Mozambique, especially in the provinces of Tete and Niassa, as well as in Zambia and Zimbabwe. In the latter, according to some estimates, it ranks as the third most widely used local language, after Shona and Ndebele. The countries of Malawi, Zambia, and Mozambique overwhelmingly constitute the central location of chiNyanja.

The language derives its name from the lake that is shared by Malawi, Mozambique, and Tanzania, with most if it as part of Malawi. The local word for a large expanse of water is *nyanja*. The people who lived along the shores of the lake and the banks of the Shire River called themselves aNyanja (lake dwellers). The Shire River flows from the southern extremity of the lake, formerly called Lake Nyasa, but now known as Lake Malawi, through southern Malawi into southern Mozambique to join the Zambezi River. The aNyanja (the singular form of which is *mNyanja*, where the prefix *m* is a syllabic nasal) spoke the language called chiNyanja (henceforth, Chinyanja).

Like most languages, Chinyanja has a number of regional dialectal variations. One of these, spoken in the hinterland of Malawi, is called chiChewa (henceforth Chichewa). This dialectal variation is the one that was spoken by the first president of Malawi, the late Dr Hastings Kamuzu Banda (cf. Watkins, 1937). The ascendancy of a Chewa to the presidency in independent Malawi had repercussions on language issues. President Banda argued that the classification of Chichewa as a dialect of Chinyanja was erroneous, deriving from unfortunate aspects of the history of missionary activity in the country, whose early activities were concentrated along the lake. Banda invoked aspects of history, plausible in some ways, to argue that Chichewa was the language of which Chinyanja was the dialectal variation (for pertinent observations, see Marwick, 1963).

The version of the history of the Chewa that Banda espoused was that the people who speak Chichewa, known as aChewa, trace their origins to a group

of people known as the Maravi (according to some Portuguese records) who migrated from the lower basin of the Congo in central Africa and eventually settled in the land mass now covered by Malawi, Zambia, and Zimbabwe. Pushed by wars, disease, and other maladies from the Congo area, the Maravi were the first group of Bantu peoples to move and settle in present day Malawi in the 16th century. Other Bantu groups such as the Tumbuka, Tonga, Yao, Lomwe, and Ngoni moved into Malawi long after the Maravi group had successfully established itself (see Kalipeni, 1996).

The Chewa were led by a powerful leader called Kalonga. He founded in Malawi what later came to be called the Maravi empire. In Malawi he established his headquarters or seat in a place called Mankhamba. Once settled, he decided to extend his influence by acquiring more land and having it settled by his subjects. To achieve those objectives he dispatched a number of his matrilineal relatives to establish settlements in various parts of the country. Among the relatives who traveled on were such chiefs as Mwase, who moved into the area called Kasungu, Kaphwiti and Lunda, who settled in the lower Shire valley. As they spread throughout the central and southern part of the country, into eastern Zambia, and into parts of Mozambique, including along the Zambezi River, their language spread too. The dispersion of Kalonga's relatives and the ensuing Chewa diaspora resulted in a proliferation of regional varieties of the language. The distinct names that the regional varieties acquired created the impression of the existence of a multiplicity of ethnic groups. Some of the groups identified themselves by making reference to significant features of their habitat.

Malawi is a country dominated by a huge lake that ranks as the third largest in Africa, after Victoria Nyanza (Lake Victoria) and Lake Tanganyika, and as the 12th largest in the world. As indicated, the word for a large expanse of water in Chichewa is *nyanja*, and the word for tall grass (savanna) is *chipeta*. The people who settled along the lakeshores and along the banks of the Shire River began to call themselves aNyanja, the lake people, and their particular variety of Chichewa was called chiNyanja, or simply Nyanja, the language of the lake people. Those who moved into the interior, the land of tall grass, were called aChipeta, the dwellers of the savanna land. These names began to obscure the nature of the relationship among the people. This was further complicated by the introduction of yet other labels. Thus, the advent of the Portuguese, entering the area from southern Africa in the 17th century, was accompanied

by the introduction of new labels. They had been in contact with such ethnic groups as the Xhosa, the Nyika, the Tchangani, etc. These referred to themselves as amaXhosa, amaNyika, amaTchangani, etc. Banda claimed that when the Portuguese encountered the Chewa living in southern Malawi and southern Mozambique, who referred to themselves as aNyanja, they referred to them as amaNyanja (see Banda, 1974–1975). Under the influence of Portuguese phonology, the sound *ny*, a palatal nasal, got velarized to *ng*. This gave rise to an ethnic group of amaNg'anja, whose language was called chiMang'anja, definitely not distinct from the Nyanja. Meanwhile, the Chewa who had settled around the southern end of Lake Malawi, spreading to the area surrounding Lake Chirwa, encountered another ethnic group, the Yao. The Yao predominated in Mozambique but had flowed into the southeast part of Malawi. The Yao word for a large expanse of water is *nyasa*, and they referred to the Nyanja people as aNyasa. The people had by then come to be grouped into aChewa, aChipeta, amaNg'anja, aNyanja, and aNyasa. The last designation appears to have contributed to British colonialists' eventual designation of the lake as Lake Nyasa, and of the country as Nyasaland. This is the name that the country had until independence in 1964, when the name of Malawi, apparently derived from 'Maravi,' was restored.

The multiplicity of labels under which the Chewa came to be identified was something that received some comment from various scholars. Thus, Young remarks about the language Nyanja that

it is the language of a people scattered over a large South-east-central African area, the aMaravi, who today live under at least six different names according to the area in which Europeans found them in the closing decades of the last century. And they were more or less on the same ground at least 300 years earlier since the Portuguese records give some of them the same names as they bear to-day. (Young, 1949: 53)

Earlier, Hetherwick had stated that

On the Shire River they are called Mang'anja, a merely local pronunciation of the word A-Nyanja. Around Lake-Shirwa they are best known by their Yao name A-Nyasa. (Hetherwick, 1901: 15)

Although Chichewa is widely spoken in Southern, eastern, and central Africa, spreading over the land-mass that includes Zambia, Malawi, Mozambique, and Zimbabwe, Greenberg does not mention it in his classification of African languages. In the works of Guthrie, Chichewa and Chimang'anja are listed as two dialect variations of Nyanja. He classifies Chichewa as belonging to zone N31b, being identified as the second dialect of the main language.

Chinyanja served as the main linguistic medium for the mass media in Malawi and was taught as a subject in educational institutions at both primary and secondary levels. In 1968, under political pressure from President Banda, a resolution was passed at the annual convention of the Malawi Congress Party, then the ruling and sole political party in Malawi, to have the name of the language officially changed to Chichewa. From that point the language Chinyanja became known as Chichewa in Malawi. Simultaneously, it was elevated to the status of national language. English remained in use as the official language.

The language policy adopted in Malawi that made Chichewa the national language contributed to the promotion of Chichewa through active educational programs, media usage, and other research activities. With the exception of work carried out within the University of Malawi, tied to contributions to, and adaptations of, advances in linguistic theory, the promotion and standardization of Chichewa were placed under the oversight of the Chichewa Board. The terms of reference of the Chichewa Board included monitoring of proper usage of Chichewa in the media, revising and updating the orthographic conventions, as well as engaging in lexicographic work. The sustained effort over many years to boost the status of Chichewa as the main language resulted in increased functional literacy in that language. Out of a population of around 11 million in Malawi, upwards of 65% have functional literacy or active command of this language.

The political directive that led to the change of name of the language in Malawi, from Chinyanja to Chichewa, did not carry over to the neighboring countries of Zambia and Mozambique. Political factors were definitely relevant. During the period when the language issue was being addressed in Malawi, political relations between Malawi and Zambia reached a nadir. This deserves comment.

Regional Politics and Language Issues

Soon after gaining independence from the United Kingdom in 1964, Malawi had a cabinet crisis. A number of radical political activists with more nationalist fervor broke ranks with Kamuzu Banda, then Prime Minister. These were among the political leaders who had invited Kamuzu Banda to return from his exile in Britain, and subsequently Ghana, to join in, and assume command of, the fight for independence. Following the granting of independence the young radicals got disillusioned with the direction taken by Kamuzu Banda's policies. The policies were seen as aiming to oppress the masses, to practically

deify Banda himself as a cult figure to whom the people were to pay homage, to undermine the efforts to promote Pan-Africanism, and to maintain the *status quo*, such that the colonialists would still enjoy the privileges that they had before independence. In brief, Banda's policies brought out the dictatorial tendencies that were eventually to characterize his rule and laid the groundwork for his tyrannical grip over the country's political development. In the revolt that resulted from the rift, the dissident politicians left the country and sought refuge in the neighboring countries of Tanzania and Zambia. This resulted in strained relations between Malawi on the one hand and those other countries on the other.

The situation became aggravated when, around 1967, there was an unsuccessful effort by a group of insurgents under the leadership of Yatuta Chisiza, a dissident politician, to unseat Kamuzu Banda militarily. The military escapade was viewed as having been perpetrated with the connivance or complicity, if not open support, of those neighboring countries. At the minimum, the insurgents seemed to have received logistical support from Zambia (for relevant details see Mchombo, 1998a, 1998b).

The worsening relations between the two countries meant that the policy about language, in this regard relating to change in the name, was not merely viewed as an issue internal to Malawi but, further, as yet another instance of Kamuzu Banda's grandiose scheme to identify the country or region with his cultural and linguistic heritage, a version of Chewa hegemony. As noted by Matiki, "Banda carried this idea of Chewa supremacy a little further by claiming that the dialect spoken by his clan is the best. This is nothing but linguacentrism" (Matiki, 1997: 529). Even if the logic or historical accuracy or factual basis for the name change were to prove impeccably sound and unimpeachable, Zambia was not ready to take hints or, worse, orders, from Malawi. A subsequent dispute concerning the proper borders of the countries, again perpetrated by the Malawi regime around the same time, merely exacerbated an already grave situation. Thus, in Zambia, as well as in Mozambique, the language has always remained as Chinyanja.

In Mozambique Chinyanja is native to 3.3% of a population numbering approximately 11.5 million. In Tete Province it is spoken by 41.7% of a population of 777 426, and it is the first language of 7.2% of the population of Niassa Province, whose population totals 506 974 (see Firmino, 1995). In Zambia, with a population of 9.1 million, Chinyanja is the first language of 16% of the population and is used and/or understood by at least 42% of the population, according to a survey conducted in 1978 (cf. Kashoki,

1978). It is one of the main languages of Zambia, ranking second after Chibemba (Bemba). In fact, out of the 9.1 million people of that country, it is estimated that 36% are Bemba, 18% Nyanja, 15% Tonga, 8% Barotze, and the remainder consisting of other ethnic groups including the Mombwe, the Tumbuka, and the Northwestern peoples (see Kalipeni, 1996). The figures show that at least six million people are fluent in Chinyanja.

The Ascendancy of Chinyanja

A recurrent joke in linguistics courses about the distinction between language and dialect is the quip that a language is 'a dialect with an army and a navy.' The rise of Chichewa in Malawi was intimately connected to the tenure of Kamuzu Banda, a Chewa, as president of Malawi. With altered political dispensation through the shift to democratic practice, and Banda's subsequent demise, Chichewa effectively lost the 'army and navy' that protected it from the status of dialect. Without formally or openly introducing a new language policy, Malawian scholars have felt it prudent to fall in line with the other countries in the region by restoring the name of Chinyanja to the language. This restoration of Chinyanja to its former status goes beyond mere efforts to promote regional linguistic harmonization. Within Malawi the national language policy adversely affected the other languages. Once again, as noted by Matiki, "the change of name [from Chinyanja to Chichewa] angered some people because there was no justification in changing the name other than the fact that President Banda was an ethnic Chewa" (1997: 527). Thus, the other ethnic groups in the country felt alienated, more so given the identification of Chichewa with political power and the relegation of their languages to relative obscurity (cf. Kishindo, 1994). The political transition to democratic practice and the departure of Kamuzu Banda from the political helm provided opportune occasion for implementation of more equitable access to political participation and the recognition of the cultural and linguistic heritage of the various segments of the nation. Kamuzu Banda's departure from the political scene was accompanied by the ascendancy of Bakili Muluzi, a Yao, as the second president of a democratic Malawi. Inevitably, the political changes witnessed shifting fortunes for Chichewa.

On Reverting to the Name of Chinyanja

Although use of the label Chichewa is likely to remain, there is systematic diminution of its former status. Thus, the Chichewa Board was subsequently

dissolved and, in 1996, the Center for Language Studies (CLS), with a broader scope of activities, was established to replace it. The center is affiliated to the University of Malawi, stressing its new mandate as a locus of research and scholarship, not as an organ of political ideology or an instrument of political hegemony. The activities carried out under the auspices of the CLS have included concerted efforts to document and provide linguistic descriptions of some of the endangered languages in Malawi.

Further, the erstwhile Department of Chichewa and Linguistics at the University of Malawi, pioneered by the present author, under political directive from President Banda to the University of Malawi to contribute to the enhancement of Chichewa, was renamed the Department of African Languages and Linguistics, thereby degrading further the profile of Chichewa.

The 'politically correct' stance of diminishing the profile of Chichewa in the university through the removal of direct reference to it in the name of the department, together with the establishment of the CLS, should not, however, be (mis)construed as indicating a real diminution in linguistic work on it. Given the historical circumstances which account for the preponderance of the available trained linguists working on Chichewa, and the headstart Chichewa was given in getting material prepared for educational and other purposes, it will continue to function in the capacity of the major language of the country. Of some significance, at least to public perception, is the recent introduction of radio news bulletins in the other languages such as Chilomwe (Lomwe), Chiyao (Yao), Chitumbuka (Tumbuka), Chitonga (Tonga), and Chisena (Malawi Sena). The news bulletins have served to increase people's awareness of these other Malawian languages, subtly contributing to the subjection of Chichewa to competition.

There are grounds for the restoration of Chinyanja as the main language. These include the literary tradition that Chinyanja has enjoyed (see Made *et al.*, 1976). The description of Chinyanja goes back to at least 1875; the first significant work can be traced to Alexander Riddell's publication in 1880 of *A grammar of Chinyanja as spoken at Lake Nyasa, with Chinyanja-English vocabulary*. This work, while not linguistically very significant, was followed in 1891 with the publication of George Henry's *A grammar of Chinyanja: a language spoken in British Central Africa on and near the shores of Lake Nyasa*. This was more comprehensive than the work by Riddell. In 1892 David Scott's *A cyclopaedic dictionary of the Mang'anja language spoken in British Central Africa* appeared, a work that was later to be revised

and enlarged by Alexander Hetherwick in 1929 as *Dictionary of the Nyanja language*. This still remains an authoritative dictionary of the language. Previously, in 1901, Hetherwick had produced *A practical manual for the Nyanja language*.

These descriptions of Chinyanja and the functional utility that the language enjoyed underscored its legitimacy to the status of a major language or lingua franca. The subsequent adoption of the colonialists' language policy, which recognized the position of Chinyanja in Central Africa, especially in Malawi, eliminated further detractions from its status. The establishment of the colonial administration in Malawi at the turn of the 20th century provided extra impetus to the promotion of Chinyanja. This led to the appearance of more works on Chinyanja and the emergence of more literary works and newspapers, such as *Msimbi*, a weekly Chinyanja newspaper that flourished for several years from 1951.

As Chinyanja gets rehabilitated, and some recent publications such as dictionaries now bear that name instead of Chichewa, the latter still remains undeniably the most familiar label for the language that is spoken and understood by more than half the population of Malawi. The reversion to the label of Chinyanja, just like the prior change to Chichewa, may lack a linguistic basis but it definitely fulfills political objectives.

Linguistic Aspects of Chinyanja

Chinyanja manifests typical aspects of the linguistic structure of Bantu languages. Its nominal system comprises a number of gender classes that are involved in the agreement patterning of the language, characteristic of Bantu languages. Thus, nominal modifiers agree with the head noun in the relevant features of gender and number, as will be illustrated below.

In its verbal structure, Chinyanja, just like other Bantu languages, displays an elaborate agglutinative structure. The verb comprises a verb root or radical, to which suffixes or extensions are added to form the verb stem (cf. Guthrie, 1962). The extensions affect argument structure or the number of expressible nominal arguments that the stem can support. The verb stem also has proclitics which encode such syntactically oriented information as negation, tense/aspect, subject and object markers, modals, conditional markers, directional markers, etc. (cf. Mchombo, 2004).

Chinyanja is also a tonal language, displaying features of lexical and grammatical tone. It has two tone levels, high (H) and low (L). Contour tones are attested but result from a combination of these tone levels, usually on long syllables (Mtenje, 1986). In its

Classification of Nouns

| | |
|------------------|-------------------|
| m-lenje 'hunter' | a-lenje 'hunters' |
| chi-soti 'hat' | zi-soti 'hats' |
| m-kóndo 'spear' | mi-kóndo 'spears' |

(1a) chi-soti ch-ángá lch-á-tsópanó
7-hat 7SM-my 7SM-ASSOC-now
chi-ja chí-ma-sangaláts-á a-lenje
7SM-REL.PRO 7SM-HABIT-please-FV 2-hunters
'that new hat of mine pleases hunters'

(1b) m-kóndó w-angá w-á-tsópanó
3-spear 3SM-my 3SM-ASSOC-now
u-ja ú-ma-sangaláts-á a-lenje
3SM-REL.PRO 3SM-HABIT-please-FV 2-hunters
'that new spear of mine pleases hunters'

(2a) m-nyamäta ‘boy’ a-nyamäta ‘boys’
 m-lenje ‘hunter’ a-lenje ‘hunters’
 m-käzi ‘woman’ a-käzi ‘women’

- | | | |
|------|--------------------------------|-----------------------------|
| (2b) | m-kóndo 'spear' | mi-kóndo 'spears' |
| | mü-nda 'garden' | mĩ-nda 'gardens' |
| | m-kángo 'lion' | mi-kángo 'lions' |
| (2c) | tsamba 'leaf' | ma-samba 'leaves' |
| | duwa 'flower' | ma-luwa 'flowers' |
| | phanga 'cave' | ma-panga 'caves' |
| (2d) | chi-sa 'nest' | zi-sa 'nests' |
| | chi-tösi 'chicken dropping' | zi-tösi 'chicken droppings' |
| | chi-pútu 'grass stubble' | zi-pútu 'grass stubble' |

These classes show part of the range of noun classification that is characteristic of Bantu languages. In the examples above, the singular forms of the first group above, dominated by nouns that denote animate things, constitutes Class 1, and its plural counterpart is Class 2. Of course, not all animate things are in this class. In fact it does also include some inanimate objects. The groups in (2b), (2c), and (2d) constitute Classes 3, 4, 5, 6, 7, and 8 respectively. The odd numbers indicate the singular forms and the even numbers their plural counterparts. This runs on to Classes 5, 6, 7, and 8. There is a further class, 1a. This class consists of nouns whose agreement patterns are those of Class 1 but whose nouns lack the *m(u)* prefix found in the Class 1 nouns. The plural of such nouns is indicated by prefixing *a* to the word. For instance, the noun *kalulu* 'hare', whose plural is *akalulu*, typifies this class. Each of these classes has a specific class marker and a specific agreement marker. Beginning with Class 2, the agreement markers are respectively *a*, *u*, *i*, *li*, *a*, *chi*, and *zi*. Class 1 has the agreement markers *mu* (or syllabic *m*), *u*, and *a*, depending on the category of the modifier. Consider the following:

- (3) m-lenje m-módzi a-na-bwéla ndí mí-kóndo
 1-hunter 1SM-one 1SM-PAST- with 4-spears
 come-FV
 ‘one hunter came with spears’

Table 1 Full range of noun classes in Chinyanja

| <i>Classes</i> | | <i>Prefixes</i> | | <i>Subject marker</i> | | <i>Object marker</i> | |
|----------------|-----------|-----------------|-----------|-----------------------|-----------|----------------------|-----------|
| <i>SG</i> | <i>PL</i> | <i>SG</i> | <i>PL</i> | <i>SG</i> | <i>PL</i> | <i>SG</i> | <i>PL</i> |
| 1 | 2 | m(u)- | a- | a- | a- | m(u) | wa |
| 3 | 4 | m(u)- | mi- | u- | i- | u | i |
| 5 | 6 | *li- | ma- | li- | a- | li | wa |
| 7 | 8 | chi- | zi- | chi- | zi- | chi | zi |
| 9 | 10 | *N- | *N- | i- | zi- | i | zi |
| 12 | 13 | ka- | ti- | ka- | ti- | ka | ti |
| 14 | 6 | u- | ma- | u- | a- | u | wa |
| 15 | | ku- | ku- | ku- | | | |
| 16 | | pa- | pa- | pa- | | | |
| 17 | | ku- | ku- | ku- | | | |
| 18 | | m(u)- | m(u)- | m(u)- | | | |

Here, the numeral *módzi* 'one' is marked with the agreement marker *m* but the verb has *a* for the subject marker. The *u* is used with demonstratives and when the segment that follows is a vowel. This seems to apply to most cases regardless of whether the vowel in question is a tense/aspect marker, an associative marker, or part of a stem, such as with possessives. Consider the following:

- (4) m-lenje w-ánú u-ja w-á
 1-hunter 1SM-your 11SM-that SM-ASSOC
 nthábwala w-a-thyol-a mi-kóndo
 10-humor 1SM-PERF-break-FV 4-spears
 'that humorous hunter of yours has broken the spears'

In this sentence, the glide *w* replaces *u* when a vowel follows regardless of the function associated with that vowel.

Although most of the nouns are bimorphemic, there are a number of cases where a further prefix, marking either diminution or augmentation, is added to an already prefixed noun. This is shown in the following:

- (5) ka-m-lenje k-ánú ka-ja k-á
 12-1-hunter 12SM-your 12SM-that 12-ASSOC
 nthábwala k-a-thyol-a ti-mi-kóndo
 10-humor 12SM-PERF-break-FV 13-4-spears
 'that small humorous hunter of yours has broken the tiny spears'

In this sentence, the preprefixes *ka* for singular and *ti* for plural get attached to nouns to convey the sense of diminutive size. These preprefixes then control the agreement patterns (cf. Bresnan and Mchombo, 1995), providing the rationale for regarding them as governing separate noun classes. One significant point to be made is that locatives also control agreement patterns. Consider the following:

- (6) ku mudzi kw-ánu kú-ma-sangaláts-á
 17-at 3-village 17SM-your 17-HABIT-please-FV
 aléndo
 2-visitors
 'your village [i.e., the location] pleases visitors'

This gives such locatives the appearance of being class markers, giving rise to the view that in Chinyanja locatives are not really prepositions that mark grammatical case but, rather, class markers (for some discussion, see Bresnan, 1991, 1994).

The full range of noun classes in Chinyanja, together with their restive subject and object markers, is given in Table 1. Some of the classes have prefixes that are starred. These classes consist of nouns that, normally, lack the indicated prefix in the noun morphology. Samples of Class 5 nouns are provided above. Most of the nouns in Classes 9 and 10 begin

with a nasal but there are no overt changes in their morphological composition that correlate with number. The number distinction is reflected in the agreement markers rather than in the overt form of the noun. Examples of Class 9/10 nouns are *nyumba* 'house(s)', *nthenga* 'feather(s)', *mphini* 'tattoo(s)', *nkhôndo* 'war'. Class 15 consists of infinitive verbs. The infinitive marker *ku-* regulates the agreement patterns, just like the diminutives (Classes 12 and 13) and locatives. The infinitives are thus regarded as constituting a separate class although, just as with the locatives, there are no nouns that are peculiar to this class. There are minor exceptions to locatives. These have to do with such words as *pansi* 'down', *kunsi* 'underneath', *panja* 'outside of a place', *kunja* '(the general) outside', *pano* 'here (at this spot)', *kuno* 'here (hereabouts)', *muno* 'in here'. With these, the locative prefixes *pa*, *ku*, and *mu* are attached to the stems *-nsi*, *-nja*, and *-no*, which are bound. The agreement pattern regulated by the infinitive marker *ku-* is exemplified by the following:

- (7) ku-imbá kw-ánu kú-ma-sangaláts-á
 15INF-sing 15SM-your 15SM-HABIT-please-FV
 a-lenje
 2-hunters
 'your singing pleases hunters'

Bibliography

- Banda H K (1974–1975). *Lectures on Chichewa at Chancellor College by His Excellency the Life President Ngwazi Dr. H. Kamuzu Banda*. Blantyre, Malawi.
- Bleek W H I (1862–1869). *A comparative grammar of South African languages*. London: Trübner.
- Bresnan J (1991). *Locative case vs. locative gender*. Paper delivered at the Berkeley Linguistics Society. Berkeley: University of California.
- Bresnan J (1994). 'Locative inversion and the architecture of universal grammar.' *Language* 70(1), 72–131.
- Bresnan J & Mchombo S A (1987). 'Topic, pronoun and agreement in Chichewa.' *Language* 63(4), 741–782.
- Bresnan J & Mchombo S A (1995). 'The lexical integrity principle: evidence from Bantu.' *Natural Language & Linguistic Theory* 13, 181–254.
- Chichewa Board (1980). *Chichewa orthography rules*. Zomba: Chichewa Board.
- Chinyanja orthography rules (1931). Zomba: The Secretariat.
- Firmino G D (1995). *Revisiting the 'language question' in post-colonial Africa: the case of Portuguese and indigenous languages in Mozambique*. Doctoral dissertation, University of California, Berkeley.
- Guthrie M (1962). 'The status of radical extensions in Bantu languages.' *Journal of African Languages* 1, 202–220.
- Hailey F (1938). *An African survey: a study of problems arising in Africa south of the Sahara*. London: Oxford University Press.

- Hetherwick A (1901). *A practical manual of the Nyanja language*. London: Christian Knowledge Society.
- Kalipeni E (1996). 'Chewa and related groups in Malawi/Zambia.' In Gall T (ed.) *Worldmark Encyclopedia of Cultures and Daily Life 1: Africa*. Pepper Pike, OH: Eastword Publications Development. 98–102.
- Kashoki M (1978). 'Between-language communication in Zambia.' In Ohannessian S & Kashoki M (eds.) *Language in Zambia*. London: International African Institute.
- Kishindo P J (1994). 'The impact of a national language on minority languages: the case of Malawi.' *Journal of Contemporary African Studies* 12(2), 127–150.
- Mackenzie D (1911). *Notes on Tumbuka syntax*. Nyasaland: Livingstonia Mission.
- Made S M, Mangoche Mbewe M V B & Jackson R (1976). *100 years of Chichewa in writing, 1875–1975*. Zomba.
- Malawi Congress Party Convention (1983). *Malawi Congress Party Convention resolutions, 1965–1983*. Blantyre, Malawi: Blantyre Print.
- Marwick M G (1963). 'History and tradition in east central Africa through the eyes of Northern Rhodesian Cewa.' *Journal of African History* 4, 375–390.
- Matiki A (1997). 'The politics of language in Malawi: a preliminary investigation.' In Herbert R K (ed.) *African linguistics at the crossroads: papers from Kwaluseni*. Cologne: Rüdiger Köppe. 521–540.
- Mchombo S (1998a). 'Democratization in Malawi: its roots and prospects.' In Gros J-G (ed.) *Democratization in late twentieth-century Africa: coping with uncertainty*. Westport, CT: Greenwood Press. 21–40.
- Mchombo S (1998b). 'National identity, democracy, and the politics of language in Malawi and Tanzania.' *Journal of African Policy Studies* 4(1), 33–46.
- Mchombo S (2004). *The syntax of Chichewa*. Cambridge: Cambridge University Press.
- Mtenje A (1980). *Aspects of Chichewa derivational phonology and syllable structure constraints*. M.A. dissertation, Southern Illinois University.
- Mtenje A D (1986). *Issues in the nonlinear phonology of Chichewa*. Doctoral dissertation, University College London.
- Price T (1940). 'Nyanja linguistic problems.' *Africa* 13, 125–137.
- Price T (1946). *The elements of Nyanja*. Blantyre: Nyasaland Church of Scotland Mission.
- Sanderson M M & Bithrey W B (1939). *An introduction to Chinyanja*. Glasgow: African Lakes Corporation. Edinburgh/London: Oliver & Boyd.
- Scott D C (1892). *A cyclopaedic dictionary of the Mang'anja language spoken in British Central Africa*. Edinburgh: Foreign Mission Committee of the Church of Scotland.
- Thomson T D (1942). *Military Chinyanja*. Zomba: Government Printer.
- Thomson T D (1955). *A practical approach to Chinyanja*. Zomba: Government Printer.
- Watkins M H (1937). *A grammar of Chichewa, A Bantu language of British Central Africa*. Philadelphia: Linguistic Society of America.
- Young C (1949). 'A review of *A practical approach to Chinyanja with English-Nyanja vocabulary* by T D Thomson, Salisbury, 1947.' *Africa* 19, 253–254.

This page intentionally left blank



Occitan

T T Field, University of Maryland, Baltimore, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Occitan is the term used today to refer to the language that evolved out of Latin in southern France. Long called ‘Provençal’ and still referred to as ‘la langue d’oc,’ Occitan is the indigenous language of a region that covers approximately a third of France, the Aran Valley of Spain, and the upper Alpine valleys of Italy.

Status

It is difficult to speak of the sociolinguistic situation of Occitan today other than in terms of marginality. The diglossia that characterized usage among rural speakers a few decades ago, when Occitan served the domains of traditional agriculture, storytelling, and the like, has given way to what might be termed ‘motivational distribution.’ In other words, for most of the population, the use of Occitan is no longer clearly tied to any particular social domain, but rather is predictable only as a function of the enthusiasm of the speaker for the language and of his or her interlocutors’ ability to manage in it. Outside the major cities, between 20 and 30% of the population claim to speak the language, though 40–50% say they understand it. This suggests that the number of speakers may be in the range of 2 million, with perhaps twice that number able to understand.

Occitan attained official status in the Aran Valley in 1983 and in Italy in 1999. In France, however, progress toward official recognition has been slow and uneven. Although the language has been present in the educational system on a limited basis since 1951, France as a whole remains committed to the anticomunitarian ideologies of the Third Republic and has refused to ratify the European Charter for Regional or Minority Languages.

Structure

The language to which Occitan is most closely related is Catalan, and it is increasingly common to

classify both as members of an Occitano-Romance group, distinct from North Gallo-Romance and Ibero-Romance proper. As in French and Catalan, Occitan lost Latin final unstressed vowels, with the exception of -a (*filh* ‘son,’ *pan* ‘bread,’ *farina* ‘flour’). Occitan phonology is distinctive historically in its failure to undergo the Romance diphthongization (*pòt* ‘he can,’ *pè* ‘foot’); in its maintenance of /aw/ (*causa* ‘thing’); and in a vowel chain shift that fronted Vulgar Latin /u/ to [y], raised /o/ and unstressed /ɔ/ to [u], and continues to raise /a/ to [o] in unstressed position (*madura* [madúro, modúro] ‘ripe (f.)’).

Occitan is a prodrop language and resembles Ibero-Romance in its morphology and syntax. However, it maintained, as did French, a two-case inflectional system into the 13th century. The most striking grammatical feature to be found in Occitan is the enunciative particle, which is limited to Gascon; it cooccurs with tense and serves discourse-level functions: *Joan que venó la vaca* ‘John [neutral assertion] sold the cow.’

Diversity

The Occitan domains never developed institutions that promoted linguistic unity, and mutual intelligibility across regions is uneven. The major dialects are Gascon, Limousin, Auvergnat, Languedocian, and Provençal. Most linguists also identify a Vivaro-Alpine dialect. Gascon, spoken in the southwest from the Garonne to the Pyrenees, is certainly the most distinctive of these in phonology, as well as in grammar, and may well deserve separate-language status within Occitano-Romance.

Standardization has been a hotly debated issue over the past few decades. Today, most activists have adopted the orthographic norms of the Institut d’Estudis Occitans; this system ensures a level of morphophonemic abstraction sufficient to allow crossdialectal comprehension. However, the Standard Occitan proposed by the Institut has not been particularly successful. It appears today that the majority of activists are ready to see Occitan as

a polycentric language, with regional norms in Gascony, Languedoc, Limousin, Provence, etc.

History

The earliest extensive Occitan texts date from the 11th century. The 12th century marks the opening of the language's classical period, when the troubadours (an Occitan word) produced their stunningly innovative poetic tradition and launched a genre dialect that would remain an international model of poetic creativity for nearly two centuries. However, as most of the Occitan regions were integrated into the kingdom of France, the language lost ground to French. In Bearn and Lower Navarre, it retained official status through the 18th century. By 1900, French had established its 'high' status in a diglossic situation that continued to evolve to its advantage. Although there were still a few children reared as monolinguals in Occitan in the early 1950s, the language had nearly disappeared from the cities and larger towns and was almost universally associated with backwardness and ignorance.

Two major movements have had the goal of revitalizing Occitan. The first of these was the 'Felibrige,' founded in 1854 and centered on the personality of Frederic Mistral. This movement had an enduring influence in Provence, and it may well account for the vitality of Provençal in the face of a very heavy influx of outsiders. The second movement, 'Occitanism,' aimed to unify the language and open up modern spaces for Occitan use. In the 1970s, this movement was responsible for a surge of public visibility for the

language and for a dramatic increase in its range of uses (e.g., theater, popular song, and academic writing). Occitanism also engendered the *Calandretas*, which are bilingual private schools in which Occitan once again plays a role in children's education and socialization.

Perspectives

Despite the progress made by activists in the final decades of the 20th century, and despite favorable official policy in small areas of Spain and Italy, Occitan continues to decline, and, as time passes, speakers who acquired the language in family and community settings are disappearing rapidly. The children who emerge as fluent speakers from the *Calandretas* and the enthusiasts who manage to pick up Occitan as a second language rarely have access to community settings in which speaking can take root. There can be no doubt that Occitanism has prolonged the life of the language and that there will continue to be people who speak, read, and write Occitan for many decades to come, but the time is near when speakers who learned the language in traditional communities will no longer exist.

Bibliography

- Bec P (1994). *La langue occitane*. Paris: Presses Universitaires de France.
 Paden W (1998). *An introduction to Old Occitan*. New York: Modern Language Association of America.

Old Church Slavonic

C M MacRobert, Oxford University, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Old Church Slavonic is the earliest Slavic literary language. It was first used in the later part of the 9th century A.D. as the vehicle of translations and original compositions by SS Cyril and Methodius and their associates for the benefit of those Slav peoples who had recently accepted Christianity. Some of these texts have survived in copies thought to date from the late 10th or 11th century, which are the primary source of information about the language and have recently been supplemented by newly discovered manuscripts; others, found in copies of later date,

can be used to provide important additional evidence about syntax and lexis.

The sound system implied by the two alphabets Glagolitic and Cyrillic, in which Old Church Slavonic was written, antedates the major change from open to closed syllable structure that took place across the Slavic languages between the 10th and 12th centuries. Some of the grammatical forms and constructions used in Old Church Slavonic manuscripts are also highly conservative, e.g., substantial remains of distinct consonantal nominal declensions, transparent postposition of the anaphoric pronoun to adjectival forms as a means of expressing definiteness, asigmatic aorist forms, and the supine with a genitive complement. The evidence of Old Church Slavonic therefore

has considerable weight in attempts to reconstruct Proto-Slavic (Common Slavonic) and to elucidate the relationship between Slavic and other Indo-European languages (see **Indo-European Languages**).

Old Church Slavonic is also the main source of information about the early history of the South-East Slavic languages (see **Bulgarian, Macedonian**). As natives of Saloniki, SS Cyril and Methodius doubtless spoke the local variety of South Slavic. As a result of their work in Moravia (863–885 A.D.), Old Church Slavonic borrowed some local items of religious terminology from Latin or Old High German, such as *mīša* < *missa*, *všodŭ* < *wizzod*, and there is some ground for supposing that at an early stage, Old Church Slavonic also incorporated certain West Slavic linguistic features, particularly in pronunciation. However, the manuscripts of South Slavic origin, from which the information about Old Church Slavonic is largely derived, preserve only traces of such a hybrid usage at best, and for the most part reflect the Slavic dialects of the southeast Balkans and the Greek terminology of the Eastern Orthodox Church.

From its inception, however, Old Church Slavonic must have differed from contemporary spoken varieties of Slavic, as it was used primarily to translate Scriptural, liturgical, and patristic texts from Greek or occasionally from Latin and Old High German. Even pronunciation may have been modified to accommodate Greek loanwords: the Glagolitic alphabet has extra letters for the velar consonants /g/ and /x/, which seem to have been reserved for use in Greek loanwords before front vowels, a position in which these phonemes did not occur in native Slavic words at that time. Comparison with the originals shows that the translations aimed at faithfulness on the basis of correspondence, phrase by phrase, between source and target. Consequently, while the grammatical forms and most of the words and semantic distinctions are Slavic, the syntax tends to mirror the constructions of the original, usually Greek.

There is, however, a range of recurrent exceptions where imitation of a foreign model would presumably have led to linguistically unacceptable results: the placing of clitics apparently follows Slavic rules; possessive adjectives or the attributive dative frequently appear in place of an attributive genitive in the original, and simple case forms may be used to translate prepositional phrases; the use of the dual number, the distribution of subordinate complementary clause, infinitive and supine, and the choices made among the elaborate past tense system of

the verb are all independent of Greek. Even the dative absolute construction, which is peculiar to Old Church Slavonic among the Slavic languages and is usually found as a translation equivalent of the Greek genitive absolute, is occasionally used to render intractable Greek constructions such as the nominalized infinitive. The compound word-formations of Greek were also frequently reproduced in Old Church Slavonic, e.g., *pravoslovie* (later *pravoslavie*) < *ορθοδοξία*. Texts believed to be original Old Church Slavonic compositions display the same type of language, which can be characterized as a compromise between early Slavic idiom and Greek literary usage in a balance so delicate that it was not subsequently maintained (see **Church Slavonic**).

Bibliography

- Auty R (1976). 'Lateinisches und althochdeutsches im altkirchenslavischen Wortschatz.' *Slovo* 25–26, 169–174.
- Aitzetmüller R (1991). *Altbulgarische Grammatik als Einführung in die slavische Sprachwissenschaft* (2nd edn.). Freiburg i. Br.: U. W. Weiher.
- Динеков П et al. (eds.) (1985–2004). *Кирило-Методиевска Енциклопедия* (4 vols). Sofia: Bălgarska Akademija na naukite.
- Huntley D (1993). 'Old Church Slavonic.' In Comrie B & Corbett G G (eds.) *The Slavonic Languages*. London/New York: Routledge. 125–187.
- Jagić V (1913). *Entstehungsgeschichte der kirchenslavischen Sprache*. Berlin: Weidmann.
- Kurz J et al. (eds.) (1966–1997). *Lexicon linguae palaeoslovenicae. Slovník jazyka staroslověnského* (5 vols). Prague: Československá Akademie Věd.
- Lunt H G (2001). *Old Church Slavonic grammar* (7th edn.). Berlin: Mouton de Gruyter.
- Sadnik L & Aitzetmüller R (1955). *Handwörterbuch zu den altkirchenslavischen Texten*. The Hague: Mouton/Heidelberg: Carl Winter.
- Schaeken J (1998). 'Palaeoslovenica. Würdigung neuentdeckter Handschriften.' In Barentsen A A et al. (eds.) *Studies in Slavic and general linguistics 24: Dutch Contributions to the 12th International Congress of Slavists*. Amsterdam/Atlanta, GA: Rodopi. 351–376.
- Schenker A M (1995). *The dawn of Slavic: an introduction to Slavic philology*. New Haven/London: Yale University Press.
- Vlasto A P (1970). *The entry of the Slavs into Christendom*. Cambridge: Cambridge University Press.
- Večerka R (1989–2002). *Altkirchenslavische (Altbulgarische) Syntax* (4 vols). Freiburg i. Br.: U. W. Weiher.

Omaha-Ponca

C Rudin, Wayne State College, Wayne, NE, USA

K Shea, University of Kansas, Lawrence, KS, USA

© 2006 Elsevier Ltd. All rights reserved.

Speakers and Linguistic Resources

Omaha-Ponca is the name linguists use for the language of the Omaha and Ponca peoples. *Umo^oho^o* (Omaha) and *Pa^oka* (Ponca, sometimes spelled Ponka) dialects differ only minimally, but are considered distinct languages by their speakers. Both tribes formerly inhabited areas near the Missouri River in northeastern Nebraska. The Omahas are still located in this area, with the tribal headquarters at Macy, Nebraska, but most of the Poncas were removed in 1878 to northern Oklahoma, around Ponca City. A smaller group of Poncas still resides in Nebraska. The Omaha-Ponca language is a member of the Dhegiha branch of Mississippi Valley Siouan, closely related to the Osage, Kansa, and Quapaw languages (for more detail on the genetic relationships of the Siouan languages, see **Siouan Languages**).

The Omaha and Ponca dialects are severely endangered, with only a few dozen elderly fluent speakers of Omaha in Nebraska and of Ponca in Oklahoma. However, many younger people have some ability to speak or understand the languages, and language classes at several schools and colleges in Nebraska and Oklahoma have had some success in promoting fluency among passive speakers and semispeakers, as well as in teaching the language to children and college students. Major linguistic resources on Omaha-Ponca include the monumental text collections of James

Owen Dorsey (1890, 1891) and Dorsey's draft grammar and slip file, the ethnographic studies by Fletcher and LaFlesche (1911) and Howard (1965), Swetland's dictionary (1991), and an unpublished grammar by Koontz (1984). Several dissertations are currently in progress.

Sounds and Spelling

Traditionally, like other Native American languages, Omaha-Ponca was not written. Independently, both tribes recently adopted nearly identical spelling systems, similar to the orthography used by Fletcher and LaFlesche, but reading and writing Omaha-Ponca are still complicated by the existence of several other orthographies. In particular, the Dorsey materials, the largest source of texts in the language, are written in an idiosyncratic orthography that uses upside-down letters for unaspirated stops, 'ç' for the dental approximant, 'q' for the voiceless velar fricative, and 'c' for the voiceless alveopalatal fricative, among other unusual symbols. Most modern linguistic writings on the language use a transcription that represents tense unaspirated stops with a double letter, nasal vowels with a hook under the letter, and alveopalatal consonants with a hachek (č, š, ž, ǰ); a slightly modified transcription known as 'NetSiouan' is used for electronic communication. This has the effect that even those who are literate in Omaha or Ponca do not have easy access to most works on the language.

In this article, the orthography adopted by current school programs is used. The phonemic inventory of Omaha-Ponca, using this system, is shown in **Table 1**.

Table 1 Phonemic inventory of Omaha-Ponca

| Sound | Labial | Dental | Alveopalatal | Velar | Laryngeal |
|-----------------------------|---|----------------|-----------------|----------------|-----------|
| Stops and affricates | | | | | |
| Voiced | b | d | j | g | |
| Voiceless | | | | | |
| Plain | p | t | ch | k | ' |
| Aspirated | p ^h | t ^h | ch ^h | k ^h | |
| Glottalized | p' | t' | | | |
| Nasals | m | n | | | |
| Fricatives | | | | | |
| Voiced | | z | zh | gh | |
| Voiceless | | | | | |
| Plain | | s | sh | x | |
| Glottalized | | s' | sh' | x' | |
| Approximants | w | th | | | h |
| Vowels | | | | | |
| Oral | i e a (o) u | | | | |
| Nasal | i ⁿ a ⁿ /o ⁿ | | | | |
| Long | (doubled letter, e.g., aa) | | | | |

Several sounds require explanation. The plain voiceless stops are lax following a fricative, tense elsewhere. Glottalized consonants, which are ejective or co-articulated with a glottal stop, are rare. The back nasal vowel is spelled *oⁿ* in Omaha and *aⁿ* in Ponca. Throughout this article, for convenience, the Omaha spelling is used. It is not entirely clear whether there is more than one phonemic back nasal vowel. Phonetic vowels varying in quality from [aⁿ] to [oⁿ] to [uⁿ] occur, but are probably allophonically conditioned. A vowel *o* is written in a few words of men's speech.

The most unusual sound in Omaha-Ponca is the consonant spelled *th*. This phoneme ranges apparently freely from [l] to a lightly articulated voiced dental fricative [ð]. Historically derived from **r*, it behaves more like a liquid than a fricative, frequently occurring in syllable-initial clusters following a voiced stop (bth, gth), for instance. Because of its similarity to the sound in the English word 'this,' it is spelled *th* in the Fletcher-LaFlesche orthography and in current educational orthographies. Other systems represent it variously as *ç* (Dorsey), *ð* (Siouanist/linguistic), or *dh* (NetSiouan).

Vowel length is distinctive in accented syllables (*náⁿ aⁿde* 'heart' vs. *náⁿde* 'inside wall'), but this contrast was not recognized by linguists until the 1990s and is still marked only sporadically in written materials. Nasality is also distinctive, but sometimes difficult to hear, especially for [i] vs. [iⁿ] adjacent to a nasal consonant or in final position. For instance, 'water' can be found written as either *ní* or *níⁿ*. A downstep pitch accent occurs on the first or second syllable of the word, and is distinctive, as in *wathát^he* 'food' and *wáthat^he* 'table', though this may turn out to correlate with vowel length. Instrumental phonetic studies of Omaha-Ponca are lacking. It would be useful to have studies of the exact quality of the various stop series, *th*, and the suprasegmental features.

Morphology

Like other Siouan languages, Omaha-Ponca has complex verbal morphology but very little elaboration of other categories. There is no grammatical class of adjectives; concepts such as 'tall' are expressed by stative verbs. Adverbs, pronouns, and demonstratives are minor, uninflected categories. Nouns, other than those derived from verbs, generally contain no inflectional morphology. The exception is vocative and inalienable possessive marking of relationship terms:

| | |
|---------------------|--|
| wi-ko ⁿ | 'my grandmother' |
| thi-ko ⁿ | 'your grandmother' |
| i-ko ⁿ | 'his/her/their grandmother (sometimes also used by men)' |

| | |
|---------------------|----------------------------------|
| ko ⁿ -ho | 'grandmother!' (male vocative) |
| ko ⁿ -ha | 'grandmother!' (female vocative) |

Definiteness is marked by a series of articles that also code animacy, proximateness, position, movement, and/or plurality of the nominal that they follow. This complex definite article system is an innovation shared with other Dhegiha languages:

| | |
|--------------------------------------|---------------------------------------|
| nu ak ^h a | 'the man (proximate)' |
| nu thi ⁿ k ^h e | 'the man (obviative animate sitting)' |
| zho ⁿ k ^h e | 'the stick (long, horizontal)' |
| zho ⁿ t ^h e | 'the wood (stacked vertically)' |

The verb is the locus of most of the grammatical information in the sentence. Besides pronominal prefixes identifying subject and object of the clause, the verb may contain prefixal instrumental, locative, dative, possessive, reflexive, *suus* (reflexive possessive), and vertitive (returning motion) markers, some of which can be obscured by phonological processes. Postverbal enclitics code plurality, negation, habitual or potential aspect, evidentiality, imperative and interrogative mode, proximateness, and other categories, some marked for person. There is no category of tense (in the following examples, the abbreviations are as follows: 1s, first-person singular; 1PL, first-person plural; AGT, agent; BEN, benefactive; REFL, reflexive; POTEN, potential; AUX, auxiliary).

| |
|--|
| a-ki-g-thize-ta-mi ⁿ k ^h e |
| 1s.AGT-BEN-REFL-get-POTEN-1s.AUX |
| 'I'll get (it) for myself'. |

Omaha-Ponca is an active-stative language, meaning that verbs take one or the other or both of two sets of pronominal prefixes, an agent set and a patient sset. The regular prefixes are given in the following example (there are also several irregular conjugations):

| | 1s | 2nd person | 3rd person | 1PL |
|----------|------------------|------------|------------|------------------|
| Agent: | a- | tha- | Ø- | o ⁿ - |
| Patient: | o ⁿ - | thi- | Ø- | wa- |

Intransitive verbs take one set or the other, depending roughly on their semantics, 'active' verbs taking the agent set as their sole argument, and 'stative' verbs taking the patient set:

| | | | |
|---|--------------------------------|--|--|
| Active verb, | gthi ⁿ 'sit': | | |
| agthi ⁿ 'I sit' | thagthi ⁿ 'you sit' | gthi ⁿ 'he/she/it/they sit' | |
| o ⁿ gthi ⁿ 'we sit' | | | |
| Stative verb, | sni 'be cold': | | |
| o ⁿ sni 'I'm cold' | thisni 'you're cold' | sni 'it's cold' | |
| wasni 'we're cold' | | | |

Transitive verbs take both an agent prefix for the subject and a patient prefix for the object, e.g., *oⁿ-thi-doⁿbai* 'we see you'. There is a portmanteau form *wi-* for first-person subject with second-person

object, and an additional patient prefix *wa-* for third-person plural or indefinite object.

Syntax

Syntactically, Omaha-Ponca is a head-marking, head-final language. Postpositions follow their nominal arguments, as in *tiútanoⁿ k^he di* ‘in the yard’ (literally, ‘yard the in’). Modal and evidential auxiliaries are at the end of the clause, after the verb, as are imperative and question particles. Determiners are the rightmost element in the nominal phrase (determiner phrase) and other noun modifiers also follow the head noun (noun + clause + possessive + article):

wat^hé tu wiwíta thoⁿ
dress blue my the
‘my blue dress’

Basic sentence word order is subject-object-verb, as in the following example (PROX, proximate; EVID, evidential):

[wahóⁿthishige ak^ha] [shóⁿgewiⁿ] [góⁿtha-i-t^he]
Orphan.Boy the horse one want-PROX-EVID
‘Orphan Boy wanted a horse.’

Full subject-object-verb (SOV) sentences are actually rather uncommon, however. All constituents except the verb are optional, so subject and/or object are often missing; a verb alone constitutes a full grammatical sentence. In addition, SOV order is far from rigid; it is not uncommon for a major constituent, such as the underlined phrase in the following example, to occur after the verb. Such postverbal phrases generally seem to be topics, but may sometimes be simply an afterthought:

M. S. izházhe athíⁿ nú ak^há
name had man the
‘The man was named M. S.’.

Because all participants are marked on the verb and all nominals are optional, it is possible to analyze Omaha-Ponca as a pronominal argument language, in the sense that the pronominal affixes on the verb are the true syntactic arguments of the clause, with nominal phrases (when they occur) being adjuncts. As in other languages, this analysis is controversial.

Relative clauses in Omaha-Ponca are internal headed, with the head noun contained within the clause. The head noun is indefinite (not marked with a definite article), whereas the clause is followed by an article appropriate to the head noun’s role in the matrix clause:

[[shiⁿnuda noⁿba uxpátheawathe] ama]
dog two I.lose.them the
‘The two dogs that I lost’.

Various types of nominal and adverbial subordinate clauses also exist, sometimes also marked with an article:

[[that^hi] t^he] úudoⁿ
you.arrive.here the good
‘It’s good that you’re here’.

Usage: Gendered Speech and Dialects

Some aspects of Omaha-Ponca language differ by the gender of the speaker. Male/female speech forms play only a minor role in the grammar and lexicon of the language; however, they are of great cultural salience and occur with high frequency, including, as they do, forms of address, greetings, terms for certain relatives, speech act markers (command, exclamation, and question particles), and interjections (see the following examples). Gendered speech sometimes hampers language teaching and revival efforts; males in particular are wary of learning inappropriate speech patterns from a female teacher. For example, *aho!*, a greeting or interjection showing approval, is used only by males. Imperative enclitics (Example (1)) and relationship terms/vocative enclitics (Example (2)) provide additional examples:

- (1) -ga (male)/-a (female), sometimes with stress shift:
oⁿi-ga/oⁿi-á ‘give it to me’ (male/female)
- (2) zhiⁿthé-ho ‘older brother!’
(male; i.e., addressed by brother)
tinu-há ‘older brother!’
(female; i.e., addressed by sister)

Differences between Omaha and Ponca varieties of the language are slight, and mostly involve recently innovated vocabulary, such as ‘telephone’ (Ponca *máⁿ aⁿze ut^hiⁿ* ‘tapping iron’ (originally ‘telegraph’) vs. Omaha *móⁿoⁿze íutha* ‘talking iron’), or ‘cup’ (Ponca *uxpé zhiⁿga* ‘little dish’ vs. Omaha *niúthatoⁿ* ‘drink water in it’). Some words differ in meaning. For instance, *shóⁿzhiⁿga* (literally ‘small horse’) means ‘colt’ in Omaha but ‘puppy’ in Ponca. *Shóⁿge* (originally ‘dog’) has shifted its meaning to ‘horse’ in both Omaha and Ponca, but the young-animal term derived from it retains its older meaning in Ponca. Such lexical differences are not necessarily absolute. Given the close contact between Omahas and Poncas, in many cases both forms may be known in both communities.

Phonological and grammatical differences between Ponca and Omaha have not been well researched. There is some indication that Ponca speakers retain the final -i of the proximate/plural, which present-day Omaha speakers drop in most environments, though ablaut shows that it is underlyingly present, as in

Ponka *athái*, Omaha *athá* 'she/he/they go', from *athé* + *i*. However, given the small number of speakers recorded, this may be more an idiolectal than a dialectal difference. In general, speakers from the two communities have no trouble understanding each other.

Bibliography

- Dorsey J O (1890). *The Cegiha language. Contributions to North American ethnology* VI. Washington, D.C: Government Printing Office.
- Dorsey J O (1891). 'Omaha and Ponka letters.' *Bureau of American Ethnology, Bulletin* 11, 1–127.
- Dorsey J O (189?). A grammar and dict[ionary] of the Ponka language prepared by the Rev. J. Owen Dorsey, etc. Manuscript, 4800 Dorsey papers, Smithsonian Institution, Washington, D.C.
- Fletcher A C & LaFlesche F (1911). 'The Omaha tribe.' *27th annual report of the Bureau of American Ethnology*. [Reprinted in two volumes, University of Nebraska Press, 1972.]
- Howard J H (1965). 'The Ponka tribe.' *Bureau of American Ethnology, Bulletin* 195. [Reprinted, University of Nebraska Press, 1995.]
- Koontz J E (1984). Preliminary sketch of the Omaha-Ponka language. Manuscript, University of Colorado, Boulder.
- Swetland M (1991). *Umo"ho" Iye of Elizabeth Stabler: a vocabulary of the Omaha language with an Omaha to English lexicon* (2nd edn.). Macy, NE: Private publication.

Omotic Languages

R J Hayward

© 2006 Elsevier Ltd. All rights reserved.

The Omotic languages constitute an indigenous Ethiopian family of the Afroasiatic phylum. They are spoken in the west and the southwest of the country, with the River Omo as a geographical locus, and from this the name derives.

Earlier opinion included Omotic within Cushitic, but subsequent to Fleming (1969), it has generally been regarded as an independent family. There is a well-founded division into North and South sub-families. While the latter (comprising Hamar, Dime, and the Ari dialects) exhibit close internal affinities, there is greater diversity within Northern Omotic, the following groups being recognized: Gongga (Kafa varieties, Mocha, Nao, and Anfillo), Dizoid (Dizi varieties and Sheko), Mao varieties, Gimira (Benchnon and She), Omoto (an extensive cluster of languages and dialects including Wolaitta, Dorze, Gamo, Gofa, Basketto, Male, Zayse, Koyra and, possibly, Chara), and Yemsa (an isolate). The greatest numbers of speakers belong to the Omoto, Gongga, and Ari groups, though accurate figures for speakers of the Omotic languages remain unrecorded.

Omotic languages exhibit many of the linguistic features typical of the area; they show especially strong typological affinities with East and Central (Agaw) Cushitic.

Syntactically: (a) They are strictly head-final, i.e., the verb is final, all nominal modifiers precede the noun, only postpositions occur, etc.; the morphology,

moreover, is entirely suffixal. (b) There is no WH-movement, though, as with any focused constituent, WH-elements may be moved to sentence-initial position by means of a type of clefting operation. (c) Verbs in non-final (but non-subordinate) clauses commonly lack agreement, and function rather like 'serial verbs.' (d) Many Omotic languages have a case system which opposes a marked nominative in subject NPs to an unmarked absolutive form found in all other syntactic functions (complement of verb, copula, or postposition) as well as for citation purposes. (e) Contrastive argument structures of lexically related verbs (e.g., passives, reciprocals, causatives, etc.) are indicated by stem suffixes. (f) Tense and modal distinctions are carried by auxiliaries following the main verb.

Phonologically: Omotic languages share the following areal features: (a) The 'emphatic' obstruent series of Afroasiatic is represented by glottalized segments. (b) There is a symmetrical system of five peripheral vowels. (c) Length is pertinent for both consonants and vowels. (d) Pitch variation functions contrastively, though functionally many of the languages probably have 'tonal accent' rather than 'paradigmatic tonal' systems.

Going beyond areal typology, a range of Omotic languages exhibit phenomena that make it plausible to hypothesize four family-specific features, which, one assumes, are inherited from the protolanguage: (a) A root-structure constraint disallowing co-occurrence of palatal ([j, ʒ, tʃ, tʃʰ, dʒ]) and non-palatal ([s, z, ts, tsʰ, (dz)]) sibilants. (b) A nasal suffix accusative marker. (c) A three-term tonal system. (d) A lexical classification of nominals in terms of vocalic suffixes.

Certain Afroasiatic features have undergone simplification in Omotic: (a) Except in the case of human animates, formal agreement for nominal gender has been neutralized. (b) Number categories and morphology have been simplified; singulative forms are relic only, and each language employs just one plural formative. (c) No trace of the Afroasiatic Prefix Conjugation has survived.

Certain (groups of) languages have developed characteristics of some interest: (a) Benchnon Gimira has a system of six tones, which makes it unique within Africa. (b) The Omoto languages have evolved a distinct series of interrogative verb paradigms employed both for Yes/No and WH-questions.

Bibliography

- Bender M L (1975). *Omotic: a new Afroasiatic language family*. Southern Illinois University Museum Series, 3. Carbondale, IL: Southern Illinois University.
- Bender M L (ed.) (1976). *The non-Semitic languages of Ethiopia*. East Lansing, MI: African Studies Center, Michigan State University.
- Fleming H C (1969). 'The classification of West Cushitic within Hamito-Semitic.' In McCall D *et al.* (eds.) *Eastern African history*. Boston University Papers on Africa. New York: Praeger.
- Hayward R J (1990). *Omotic language studies*. London: School of Oriental and African Studies.

Oneida

C Abbott, University of Wisconsin, Green Bay, WI, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Oneida is a Native American language of the northern branch of the Iroquoian family, related to Seneca, Cayuga, Tuscarora, Onondaga, and, most closely, Mohawk. The homeland of the Oneida people is in central New York state. Migrations in the 1800s led to the three current communities of Oneidas on reservations in central New York and near Green Bay, Wisconsin, and a reserve near London, Ontario. A century ago most Oneidas spoke the language, but currently all Oneidas speak English and there are only small numbers of Oneida native speakers, primarily in Wisconsin and Ontario. All three Oneida communities sponsor efforts to preserve the language, but it is definitely endangered with the total number of fluent native speakers under 100. The Oneidas' name for themselves is *onayote?a'ká* 'people of the standing stone.' They also use the term *ukwehu' wé* 'native people' for themselves and other Iroquoian people.

The oral traditions of the Oneidas support a wealth of stories and a rich set of ceremonies, shared with other members of the League of the Iroquois (also known as the Six Nations or *Haudenosaunee*). A written form of the language is a recent innovation. Jesuit missionaries established a writing tradition for the Mohawk language, and it was used by a few people for Oneida in the 19th century for personal letters, some records, and Bible translations.

A linguistically based orthography was invented in the late 1930s and, slightly revised, has been in use since the 1970s for many language preservation materials.

Phonology

The Oneida phonemes are four oral vowels /i, e, a, o/, two nasal vowels /ʌ, u/, four resonants /l, w, y, n/, two stops /t, k/, a fricative /s/, two laryngeals /h, ʔ/, and a phoneme of vowel length. Two affricates are often analyzed as phoneme combinations /tsy, tshy/. The voicing of the stops and palatalization of the fricative are subphonemic processes conditioned by the following sound. Vowel length and pitch are distinctive, and the prosodic patterns produce one of the principal contrasts with the related Mohawk language. Patterns of epenthesis are another important contrast.

The sound system of the language is remarkable for a number of features: the small inventory of phonemes, the lack of labial sounds, and the presence of whispered syllables in a morphophonological process conditioned by placement within sentences. For example, the word for 'sugar' is *onutákli?* when it is followed by other words and *onutákehli* with the last syllable *-li-* whispered when it occurs at the end of a sentence or before a major phrase.

Morphology and the Lexicon

The morphology of the language is complex. There are only three clear word classes (nouns, verbs, and particles), but affixation is common with nouns and especially with verbs, which require, at minimum,

pronominal prefixes and aspectual suffixes added to either simple or complex verb stems. There is a rich set of derivational processes that manipulate the basic argument structure of verbs. The process of noun incorporation, along with derivational morphemes such as reflexives, benefactives, causatives, and instrumentals, can build complex stems from simpler roots. There are also devices that convert nouns to verbs and verbs to nouns. Thus, derived forms can nest within others to create words of amazing length. In addition, there is a rich set of inflectional affixes for verbs. Suffixes supply aspectual and some tense inflections. One set of prefixes supplies a pronominal coding of one or two arguments (agent and patient) with number, person, and gender distinctions. There are two distinct feminine genders along with a masculine gender and a neuter gender that largely overlaps with one of the feminine genders. The other feminine gender is the unmarked gender in the singular, whereas the masculine is the unmarked gender in the plural. There are three categories of number: singular, dual, and plural. In addition verbs may have up to six of an additional set of 11 prefixes that supply various adverbial, directional, tense, mood, lexical, and syntactic functions. Words thus contain quite a few morphemes, and these morphemes are subject to quite a bit of alternation, conditioned by surrounding morphemes, by surrounding phonemes, and by accentuation patterns.

A couple of examples of Oneida verb forms demonstrate the template: Prefixes-Pronominals-Verb stem-Aspect suffix.

- (1) t -á -t -k -e -ʔ
 back -will -toward -I -go -ASP
 'I will come back'

In (1), the verb stem *-e-* 'go' has a punctual aspect suffix *-ʔ-* and a pronominal prefix *-k-* that indicates 'first-person singular agent.' The sequence *tat* is a combination of three prefixes: *-t-* 'direction toward,' *-á-* 'future tense,' and *-t-* 'returning.'

- (2) t -huwati -lihunyani -haʔ
 there -they.AGT/them.PAT -teach -HABIT
 'school, they teach them there'

The form in (2) is constructed as a verb, but can function as a noun. It consists of a complex verb stem *-lihunyaniht-*, which is made of simpler components: an incorporated noun, *-lihw-* 'custom'; a verb root, *-uni-* 'make'; a benefactive derivational form that allows an argument role in the pronominal prefix for the receiver of the teaching, *-ani-*; and an instrumental derivational suffix that allows a focus on the means of teaching, in this case, the location,

-ht-. When these four components are combined, certain sound rules apply: *-w-* is lost before *-u-*, *-i-* becomes a consonant before a vowel, and *-h-* is lost in the *-hth-* combination. The aspectual suffix *-haʔ* is 'habitual,' the pronominal prefix *-huwati-* indicates a third-person plural agent and patient, and the initial prefix *-t-* indicates location.

The noun morphology is simpler. A few nouns are uninflected, but most nouns have obligatory prefixes and suffixes on basic noun roots, and these affixes mark the resulting words as nouns. The basic nominal prefixes can be replaced by possessive prefixes. There are a variety of locative suffixes, several pluralizing suffixes, a number of verb roots that function as adjective suffixes, and a few other suffixes. Many words that function as nouns in sentences, however, are not built from noun roots but instead are verb forms that produce descriptions used as nominals. In a few cases, it is difficult to tell whether a word is a noun or a verb.

- (3) ka -lút -eʔ
 it -log -NOM
 'log'
- (4) ka -lut -o'kú
 it -log -under
 'under the log'
- (5) ka -lu't -ót -eʔ
 it -log -stand -ASP
 'tree, standing log, the tree is standing'

The prefix *ka-* is both a common noun prefix and a neuter pronominal prefix for verbs, and the suffix *-eʔ* is both a noun suffix and a verb suffix for the stative aspect.

As a result of the complex morphology, Oneida provides its speakers with enormous resources for word building. Undoubtedly not all of this potential is exploited, but many forms that are used are lexicalized, often with some semantic specialization, as in the verb for 'they teach them there,' which lexicalizes as 'school.' That lexicalization is sometimes marked by particular additional suffixes, but for many words there is no formal marking, only use, to indicate the lexicalization. Many functional nouns are thus created from formal verbs when a description becomes a name. This may account in part for the resistance the language has to borrowing words from English, which has surrounded and endangered the language for centuries.

Syntax

In the syntax, word order is not particularly rigid and intuitions of sentencehood are not strong. Particles

and clusters of particles connect strings of predication and sometimes link to nominal arguments (which may appear formally to be verbs). The main arguments of the predication are encoded into the verb in the pronominal prefixes, and, if there is need to elaborate them, any elaboration tends to follow the predication. Particles and combinations of them provide discourse functions, subordination markers, deictics, emphasis, evidentials, and the pacing devices that are well developed in the oral tradition.

Scholarship

The academic study of the language includes some early text collection by Boas (1909) and analysis of verb stem classes by Barbeau (1915), but the real foundational work is by Lounsbury, based on fieldwork done in the Wisconsin community in the late 1930s and early 1940s. His work (Lounsbury, 1953), based on his M.A. thesis on Oneida phonology and his doctoral dissertation on verb morphology, set a framework not only for the future study of Oneida but for all the northern Iroquoian languages. Subsequent work by Karin Michelson has advanced the understanding of the sound system (Michelson, 1988) and the aspectual system (Michelson, 1995). The lexicon of the language is documented in two dictionaries, one based on fieldwork from the Wisconsin community (Abbott *et al.*, 1996) and one from the Ontario community (Michelson and Doxtator, 2002). A sketch of the linguistic structure of the language is available in Abbott (2000), and more complete grammars, both for reference and teaching, are in preparation. There are several text collections with linguistic analysis (Campisi and Christjohn, 1980; Abbott *et al.*, 1980; Michelson, 1981; Elm and Antone, 2000).

Community Work

Each of the three Oneida communities has a language preservation/recovery program to combat the endangered status of the language. Samples of the language, both written and spoken, are available on the websites of two of the Oneida communities: the Wisconsin community and the New York community. The language is being taught in tribal schools and in informal community classes. For the most part, the communities have adopted, since the 1970s, the writing system developed by Lounsbury, slightly modified from Lounsbury (1953). These programs have produced pedagogical materials, including some text collections (Abbott, 1982a, 1982b, 1983a, 1983b; Hinton, 1996) and word lists (Anton *et al.*, 1981; Anton, 1982), and include language material on their

websites. The success of these programs in stemming the language loss over the last several decades has been fairly modest.

Bibliography

- Abbott C (1981). 'Here and there in Oneida.' *International Journal of American Linguistics* 47, 50–57.
- Abbott C (1982a). *The bear stories Oneida*. Oneida, WI: Oneida Nation of Wisconsin.
- Abbott C (1982b). *Ukwehuwehnéha Onúhkwahṭ (Oneida medicine)*. Oneida, WI: Oneida Nation of Wisconsin.
- Abbott C (1983a). *Animal fables*. Oneida, WI: Oneida Nation of Wisconsin.
- Abbott C (1983b). *Witch stories*. Oneida, WI: Oneida Nation of Wisconsin.
- Abbott C (1984). 'Two feminine genders in Oneida.' *Anthropological Linguistics* 26(2), 125–137.
- Abbott C (1998). 'Lessons from an Oneida dictionary.' *Dictionaries – Journal of the Dictionary Society of North America* 19, 124–134.
- Abbott C (2000). *Languages of the world series 301: Oneida*. Muenchen: LINCOM EUROPA.
- Abbott C, Cornelius M & Johns L (1980). 'Two stories: Oneida.' *International Journal of American Linguistics – Native American Text Series* 4, 67–76.
- Abbott C, Christjohn A & Hinton M (1996). *An Oneida dictionary*. Oneida, WI: Oneida Nation of Wisconsin.
- Anton A, Doxtator M *et al.* (1981). *Tekalihwathé'tha?*. London, ON, Canada: University of Western Ontario Centre for the Research and Teaching of Canadian Native Languages.
- Anton E (1982). *Honayote'a'ká' Kawanakalatatú*. London, ON, Canada: University of Western Ontario Centre for the Research and Teaching of Canadian Native Languages.
- Boas F (1909). 'Notes on the Iroquoian language.' In *Putnam anniversary volume: anthropological essays presented to Frederic Ward Putnam in honor of his seventieth birthday, April 16, 1909 by his friends and associates*. Stechert: New York. 427–460.
- Barbeau C M (1915). *Geological survey, memoir 46, anthropological series 7: Classification of Iroquoian radicals with subjective pronominal prefixes*. Ottawa, Canada: Canada Department of Mines.
- Basehart H W (1953). Historical changes in the kinship system of the Oneida Indians. Ph.D. diss., Harvard University.
- Campisi J & Christjohn R (1980). 'Two letters: Oneida.' *International Journal of American Linguistics – Native American Text Series* 4, 41–44.
- Elm D & Antone H (2000). *The Oneida creation story*. Lounsbury F & Gick G (trans. and eds.). Lincoln, NE: University of Nebraska Press.
- Hinton M (1996). *A collection of Oneida stories*. Oneida, WI: Oneida Nation of Wisconsin.
- Lounsbury F (1946). Phonology of the Oneida language. M.A. thesis, University of Wisconsin.

- Lounsbury F (1953). *Oneida verb morphology*. New Haven, CT: Yale University Press.
- Lounsbury F (1978). 'Iroquoian languages.' In Sturtevant W (ed.) *Handbook of North American Indians*, vol. 15. Washington, DC: Smithsonian. 334–343.
- Michelson K (1981). *Mercury series 73: Three stories in Oneida*. Ottawa, Canada: National Museum of Man.
- Michelson K (1988). *A comparative study of Lake-Iroquoian accent*. Dordrecht: Kluwer Academic.
- Michelson K (1990). 'The Oneida lexicon.' *Proceedings of the Berkeley Linguistics Society* 16, 73–84.
- Michelson K (1991). 'Possessor stranding in Oneida.' *Linguistic Inquiry* 22, 756–761.
- Michelson K (1995). Aspect inflections of Oneida manner-of-motion verbs. Paper presented at the Conference on Iroquois Research.
- Michelson K & Doxtator M (2002). *Oneida-English English-Oneida dictionary*. Toronto, Canada: University of Toronto Press.
- Swadesh M, Lounsbury F & Archiquette O (1965). *Onaṭo-taʔa:gá· deyeliḥwahgwá·ta (Oneida hymnal)*. Oneida, WI: Oneida Nation of Wisconsin.
- ## Relevant Websites
- <http://language.oneidanation.org> – Oneida Wisconsin community.
- <http://www.oneida-nation.net> – Oneida New York community.

Relevant Websites

<http://language.oneidanation.org> – Oneida Wisconsin community.

<http://www.oneida-nation.net> – Oneida New York community.

Oromo

D Appleyard, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Oromo (self-name Afaan Oromo ‘language of the Oromo’) is one of the major languages of the Horn of Africa, spoken predominantly in Ethiopia, but also in northern and eastern Kenya and a little in southern Somalia. Estimates of numbers of speakers vary widely from about 17 300 000 (based on current Ethnologue figures) to ‘approximately 30 million’ (Griegenow-Mewis, 2001: 9), and there are probably about 2 million more who use it as a second language. Oromo is the major member of the Oromoid subgroup of the Lowland East Cushitic branch of Cushitic languages. There is currently no agreed-upon standard form of Oromo. Since it was adopted as a national language within the Oromo region in 1992, the Central-Western variety, which has the largest number of speakers, has tended to form the basis upon which a standardized form is being built. There are three main dialect clusters of Oromo, the Central-Western group, with at least 9 million speakers, comprising the Macha, Tuulamaa, Wallo, and Raya varieties, all spoken within Ethiopia; the Eastern group, also known as Harar Oromo or Qottu, spoken in eastern Ethiopia; and the Southern group, including Booranaa, Guji, Arsi, and Gabra, spoken in southern Ethiopia and adjacent parts of Kenya. Distinct from this last group are Orma, spoken along the Tana River in Kenya and apparently in southern Somalia along the Juba river, and Waata, spoken along the Kenyan coast to the south of Orma.

Under the Ethiopian imperial regime, which fell in 1974, the status of Oromo in Ethiopia was that of a spoken, vernacular language only. Its use in schools, the media, and other public forums was in effect proscribed, although Amharicized Oromos had been influential in the government of Ethiopia since the middle of the 18th century. With the advent of the Marxist regime, which gave some official recognition to Ethiopia's rich multilingual situation, Oromo was designated as one of the eventual 15 languages of the literacy campaign in Ethiopia, and printed and broadcast materials in Oromo started to appear. At first Oromo was written in a slightly adapted form of the Ethiopian syllabary, which had hitherto been mostly used for Ge'ez, or Classical Ethiopic, Amharic, and Tigrinya, but the move to write the language in Roman script, known as qubee in Oromo, soon prevailed. The decision of the Oromo Liberation Front to adopt the Roman script as early as 1974 doubtless gave this move impetus, though until 1991–1992 it was the refugee or exile community that made use of qubee. Additionally, there was at first no consensus on the representation of particular phonemes, and even today there can still be hesitations in marking vowel length.

Phonology

Oromo has 24 consonant phonemes, represented in the qubee orthography as follows, with IPA values where different between slashes as shown in **Table 1**. In addition, p, v, and z occur in loanwords, and some dialects, e.g., Eastern Oromo, also have a voiceless velar fricative /x/, typically in place of /k/ in other

Table 1 The consonant phonemes of Oromo

| | Bilabial | Alveolar/ dental | Palatal | Velar | Glottal |
|--------------------------------------|------------|---------------------|-------------------|-----------|---------|
| Plosive | b | d t | j ch /dʒ/ /tʃ/ | k g | ʔ |
| Glottalized Plosive/ affricate | ph /pʰ/ | x /tʰ/ | c /tʃʰ/ | q /kʰ/ | |
| Implosive | | dh /d/ | | | |
| Fricative | f | s | sh /ʃ/ | | h |
| Nasal | m | n | ny /ɲ/ | | |
| Lateral | | l | | | |
| Approximant | w | r | y /j/ | | |

dialects. All consonants except for ʔ and h may occur both long and short, though the orthography does not indicate long consonants where the symbol used is a digraph: eenyu /ʔe:ɲɲu [ʔ] / 'who?', buuphaa /bu:pʰa [ʔ] / 'egg'.

There are five vowels: a, e, i, o, u, each of which occurs both short and long, long vowels normally being indicated by doubling. In prepausal position, final long vowels are shortened somewhat and are closed by a glottal stop. According to dialect, in the same position final short vowels also are either closed by a glottal stop or devoiced. Some morphological clitics also cause a change in vowel length when added to vowel-final stems: nama 'the man' but namaa fi farda 'the man and the horse'. Additionally, several descriptions of Oromo dialects mention vowel length dissimilation, whereby long vowels in more than two consecutive syllables are not permitted: ijoollee + dhaaf > ijoolledhaaf 'for the children'. It has also been noted (Owens, 1985: 16) that only one long vowel per morpheme is permitted.

Consonant clusters in Oromo are limited to two components. Across morpheme boundaries, various patterns of consonant assimilation occur: dhug- + -ti > dhugdi 'she drinks', nyaadh- + -na > nyaanna 'we eat', dhaq- + -te > dhaqxe 'she went', gal- + -ne > galle 'we entered', Oromot[a]- + -ni > Oromoonni 'the Oromos' (subject case). Spoken forms of Oromo also seem to make greater use of consonant assimilation than the written language. Potential clusters of more than two consonants are always resolved by insertion of an epenthetic vowel, usually i: kenn- + -te > kennite 'you gave'. Sometimes, metathesis of the component consonants and an epenthetic vowel is also involved: arg- + -te > agarte 'you saw' beside argite.

Oromo is a tone-accent language, but details do differ somewhat from one spoken dialect to another. There is generally a simple, two-way contrast between high and non-high. As in a number of other Lowland East Cushitic languages, tone does not, however, distinguish lexical items, but is linked with morphological or syntactic categories, as for instance in Eastern Oromo xeesúmaa [L.H.H] 'the guest' in the absolute case or 'basic' form, but xeesúmaa [L.H.L] before a clitic such as the dative marker -f, or the same in sentence-final predicate position, and xeesu-maa [L.L.L] optionally before a phrase-final adjective. Written Oromo, however, does not mark accent. Interestingly, potential confusion between two particles, predicate focusing hín, with high tone, and present negative marker hin, with low tone, is avoided in written Oromo by adopting an Eastern Oromo dialect variant ni in the former sense and keeping hin as the negative marker.

Morphology

Oromo has a moderately complex morphology, both inflectional and derivational, similar in categories and extent to other Cushitic languages. Nouns show gender, number, and case marking, though the first of these is more typically detectable only in agreement rather than being formally marked on the noun. Derived adjectives, however, do mostly show gender: diimaa (masc.): diimtuu (fem.) 'red'. There are two genders, masculine and feminine, and two numbers, singular and plural. In addition, there are some singulative or particulative forms: nama 'man': namicha 'a particular man', jaarti 'old woman': jaartittii 'a particular old woman'. There are two fundamental cases, the absolute and the nominative, which generally require agreement among constituents of the noun phrase. Other case functions are only marked phrase-finally and do not elicit agreement. As with many other Cushitic languages, in Oromo the nominative or subject case is the marked form, and the absolute case, with functions ranging from predicative, direct object, and pre-clitic position to citation form, is unmarked.

abbaa-n koo nama dheeraa dha
father.SUBJ my man.ABS tall.ABS COP
'my father is a tall man'

meeshaa sana arg-ite
thing.ABS that.ABS see-2SING.PAST
'did you see that thing?'

mana keessa seen-e
house.ABS inside enter-3MASC.PAST
'he went into the house'

The nominative or subject case is formed by a range of suffixes, -n, -ni, -i, or Ø (but with tonal difference), or -ti (some feminine nouns only) added according to the shape of the absolute form.

nam-ni dureess-i asi jir-a
man-SUBJ rich-SUBJ here exist-3MASC.PRES
 'the rich man is here'

saree-n adii-n ni iyy-iti
dog-SUBJ white-SUBJ FOCUS bark-3FEM.PRES
 'the white dog is barking'

bishaan hin dhug-aam-e /biʃa:n/ [L.H]
water-(SUBJ) NEG drink-PASS-3MASC.PAST
 'the water wasn't drunk'

bishaan dhug-ani /biʃa:n/ [L.L]
water-(ABS) drink-3PL.PAST
 'they drank the water'

The remaining case functions all are built on the absolutive, either by means of clitics, both postpositions and occasionally prepositions, or by minor modification in the instance of the possessive case form. Possessive marking occurs only phrase-finally and is typically formed by lengthening a final vowel usually with high tone. Optionally a possessive linking particle may also be used before the possessive noun or phrase: kan with masculine head nouns, tan with feminine.

mana nam-ichaa-n beek-a
house.ABS man-SINGULATIVE.POSS-I know-1SING.PRES
 'I know the man's house'

farda kan
horse.ABS PART.MASC
 nam-ichaa arg-ite
man-SINGULATIVE.POSS see-2SING.PAST
 'you saw the man's horse'

Verbs in Oromo inflect for tense-mood-aspect (TMA), person (including gender and number as appropriate), and voice. Verb inflection is by means of suffixes, and the usual morpheme string is root + [voice] + person + TMA. The verb form may also be preceded by various proclitics or pre-verbs, such as negative, optative, and predicate focus markers, and may also have added in final position a conjunctive suffix:

loon ni bit-achi-siif-tanii-ti ...
cattle FOCUS buy-AUTOBENEFACTIVE-CAUS-2PL.
PAST-and
 'you made (someone) buy cattle for themselves and ...'

There are four main voices or derived stems of the verb in addition to the basic form: autobenefactive (sometimes also referred to as middle voice), causative, passive, and intensive or frequentative. The first

three of these are formed by suffixes, which in the instance of the causative show some considerable variation according both to the shape of the stem to which it is added and to the shape of the following personal marker. The frequentative stem is formed by means of partial reduplication of the basic stem:

cab-uu 'to break'
break-INF
 caccab-uu 'to break into pieces'
break.INTENSIVE-INF
 deebi'-uu 'to return'
return-INF
 deddeebi'-uu 'to keep on repeating'
return.INTENSIVE-INF

Up to two derived stem formatives may also be added to the basic verb stem according to prescribed sequences and combinations:

deebi'-uu 'to return'
return-INF
 deebi-s-uu 'to answer'
return-CAUSE-INF
 deebi-f-am-uu 'to be answered'
return-CAUSE-PASS-INF
 deebi-f-ach-uu 'to return s.th. for oneself'
return-CAUS-AUTOBENEFACTIVE-INF
 deddeebi-s-uu 'to repeat often'
return.INTENSIVE-CAUS-INF

Person markers, which follow the verbal stem, show the typical Cushitic pattern in which the 1st singular and 3rd masculine are formally identical, though in written Oromo the former is distinguished by suffixing to the preceding word -n, evidently a reduced form of the independent pronoun ani 'I':

mana barumsaa-n deem-a
house learning.POSS-I go-1SING.PRES
 'I go to school'
 mana barumsaa deem-a
house learning.POSS go-3MASC-PRES
 'he goes to school'

In keeping with the same underlying Cushitic pattern, the 2nd sing., and 3rd fem., also have identical personal markers, -t-, though in Oromo there is a difference of TMA vocalization in the present tense. There are three basic finite TMA paradigms: the present, the past, and what has been called the subordinate, or sometimes the subjunctive. The present is a main clause form only, while the past is employed both in main and dependent clauses. The subordinate/subjunctive form is used in a range of functions, both in dependent clauses, but also as a negative present with the particle hin, and as a jussive with the particle haa. The negative past and the negative jussive are both, on the other hand, invariable with respect to person. TMA marking is by means of

the vocalic elements following the person marker, essentially -e in the Past and -u in the subordinate/subjunctive, and -a in the Present except in the 3rd feminine, where it is -i. The 2nd and 3rd plural forms in written Oromo have the endings -tan [i] and -an [i] in all tenses, though -tu and -u also occur in the present tense and subordinate/subjunctive forms. A number of compound tenses also occur, combining variously finite tenses or verbal nouns, such as the infinitive or participle, with finite forms of such verbs as jiruu 'to be' or its Past tense equivalent turuu, or ta'uu 'to become'.

An interesting type of verb compound or composite, which has parallels across the Ethiopian language area, involves a fixed particle, typically underivable, and the verb jechuu 'to say', or in a causative-transitive function, gochuu 'to make':

| | | | |
|-----------------------------|-------|------------|------------|
| nam-ich-i | cal | jedh-ee | tur-e |
| man- | 'cal' | say- | be.past- |
| SINGULATIVE-SUBJ | | 3MASC.PAST | 3MASC.PAST |
| (cal jechuu 'to be quiet') | | | |
| 'the man was keeping quiet' | | | |

The normal word order in Oromo is SOV, as can be seen from various examples above, and dependent clauses generally precede the main clause. Relative clauses, however, follow their head noun. Subordinating particles or conjunctions are usually placed at the beginning of the clause, but may also be placed immediately before the verb at the end of the clause. Some conjunctions are disjunct, comprising both an element at the beginning of the clause and a clitic or affix placed after the verb (e.g., waan ... -f 'because' below).

| | | | | |
|-----------------|-----------------|-------------------------|---------------------|---------|
| gurbaa-n | osoo | loon | tiks-uu | waan |
| boy-SUBJ | while | cattle | watch-3MASC-SUBJUNC | because |
| midhaan | namaa | nyaach-is-ee-f | | |
| grain | man.POSS | eat-CAUS-3MAS.PAST-PART | | |
| abbaa-n-saa | reeb-ee | kur- | | |
| father-SUBJ-his | beat-3MASC.PAST | be.PAST-3MASC.PAST | | |

'because the boy, while watching the cattle, had let them eat someone else's grain, his father had beaten him'

An interesting syntactic feature that Oromo shares with most other Cushitic, and especially Lowland East Cushitic languages, is a system of focus marking by means of clitic particles with different markers for predicate and non-predicate focus. Oromo has essentially two focus constructions, both of which are optional, one used exclusively for subject focus and one for predicate focus. A third clitic is used for emphasizing non-subject nominals and is perhaps on the way to becoming a third focus marker (Griefenow-Mewis, 2001: 55). The subject focus marker is -tu[u] which is added to the absolutive case and neutralizes person/number agreement with the verb, which remains in the 3rd masculine:

| | | | |
|---------------|--------|------------|------------|
| mukk-een-tu | oddoo | keessa-tti | arg-am-a |
| tree-PL-FOCUS | garden | inside-LOC | see-PASS- |
| | | | 3MASC.PRES |

'trees can be seen in the garden'

Bibliography

- Gragg G (1976). 'Oromo.' In Bender M L (ed.) *The Non-Semitic languages of Ethiopia*. East Lansing, MI: African Studies Center, Michigan State University.
- Gragg G (1982). *Oromo dictionary*. East Lansing, MI: African Studies Center, Michigan State University, in cooperation with Oriental Institute, University of Chicago.
- Griefenow-Mewis C (2001). *A grammatical sketch of written Oromo*. Köln: Köppe Verlag.
- Griefenow-Mewis C & Bitima T (1994). *Lehrbuch des Oromo*. Köln: Köppe Verlag.
- Owens J (1985). *A grammar of Harar Oromo (Northeastern Ethiopia)*. Hamburg: Buske Verlag.
- Tucho Y, Zorc R D & Barna E C (1996). *Oromo newspaper reader, grammar sketch and lexicon*. Kensington: Dunwoody Press.

Ossetic

M Job and R Schäfer, University of Göttingen, Göttingen, Germany

© 2006 Elsevier Ltd. All rights reserved.

Ethnography, History, and Literature

Ossetic (also 'Ossetian', ISO639: 'oss') is an Iranian language spoken by approximately 650 000 people,

mainly in the Republic of North Ossetia-Alania (Russian Federation), the South Ossetic Region in Georgia, in various other parts of the Russian Federation, and in scattered settlements in Turkey. The capital of North Ossetia is Vladikavkaz (Dzæu-džiqæu in Ossetic). All speakers are bilingual (with Russian, Georgian, or Turkish as a second language) (Figure 1).

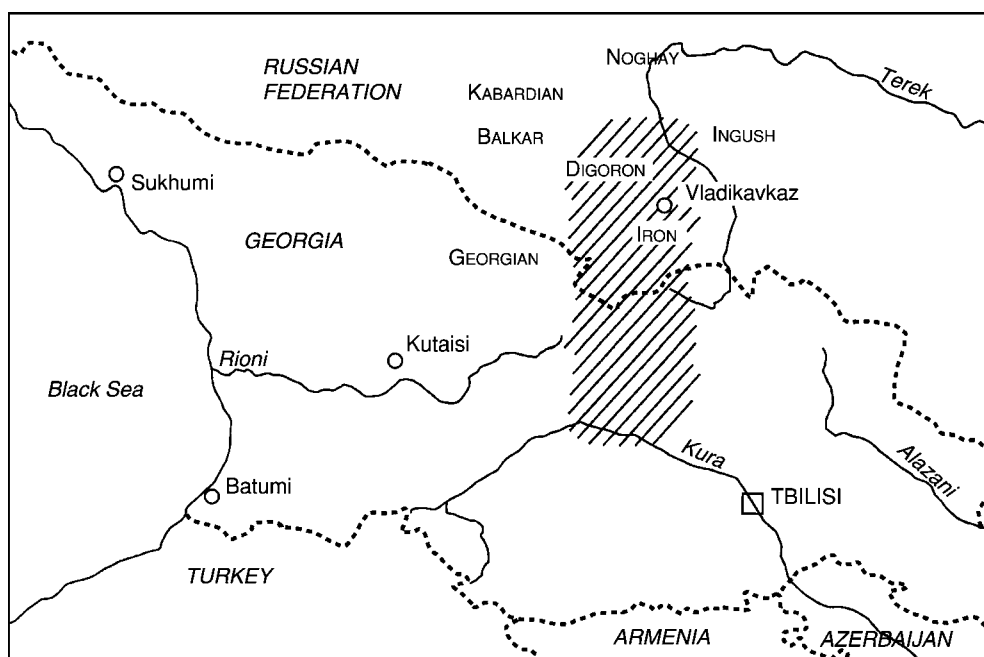


Figure 1 Ossetic area (hatched, adjacent languages in small caps).

Ossetic belongs to the Eastern Iranian branch of Indo-European of which the oldest historic member is Avestan. In the Middle-Iranian period, the Alanic group of languages comprised the closest relatives of the unattested predecessor of Ossetic. These quite fragmentarily attested languages were spoken from approximately 400 B.C. (earliest mention of the Sarmatians) to the 13th century A.D. in Southern Russia and on the Northern coast of the Black Sea. The first Ossetic document was a catechism printed in Moscow in 1798. Several writing systems based on the Georgian, Roman, and Cyrillic alphabets had been in use before Cyrillic was made official in 1939. In this article, we use the transliteration used by most scholars. The first grammatical description of Ossetic was Andreas Sjögren's 'Iron Ævzagaxur' (St. Petersburg 1844).

The two main dialects, Iron and Digoron, show some major phonological and morphological differences. Still, we will only discuss Iron, which is the basis for the literary language.

The mythological Nart tales, traditionally told by wandering minstrels, were collected from oral sources in the early 20th century by Vsevolod Miller. They have become the national epic. Its first translation into a Western language (French) was done by Georges Dumézil in 1930. Ossetic artistic poetry developed during the 19th century and found its heyday in the works of the national poet Xetægkaty K'osta (1859–1939).

Consonants

Ossetic shows a systematic opposition of voiceless aspirated, voiced, and voiceless ejective stops and affricates. The voiceless uvular stop has no ejective nor voiced counterparts.

The alveolar affricates $/ts\ dz\ tʃ/$ are realized as fricatives in Iron, except for the ejective ($[s\ z\ tʃʰ]$) and when geminated $[tsː\ dzː\ tʃː]$. In all positions, the dentoalveolar fricatives $/s\ z/$ are realized as postalveolars $[ʃ\ ʒ]$. These changes are not reflected in the orthography. An older stage is attested in Ossetic dialects in Turkey, where $/s\ z/$ are $[ʃ\ ʒ]$, but $/ts\ dz\ tʃ/$ are still $[ts^h\ dz\ tʃʰ]$ (Table 1). $[h]$, written γ , occurs in some interjections like $\gamma\ae j$ $[h\ae j]$ 'hey'.

The postalveolar affricates $[tʃ^h]$, $[dʒ]$, $[tʃʰ]$ are assimilated variants of the velars before front vowels, e.g., *kark* 'hen' and *karč-y* 'of the hen' (genitive). The few exceptions are loan-words, such as *džauyr* 'non-believer' from Circassian *džauyr*. The only regular blocking of this assimilation occurs with the superessive marker *-yl*: *kark-yl* 'on the hen'.

Since the sequence Consonant + $/ui/$ + Consonant is not licensed otherwise in Ossetic, we assume labialized stops in words like *quyn* to be phonemic: $/q^w\text{in}/$ 'hair'. Biphonemic geminated stops and affricates (which are voiceless and unaspirated) occur in lexical entries (*læppu* $[l\ae pːu]$ 'boy') or at morpheme boundaries: *dard* 'far' becomes *dard-dær* $[dardː\ r\ae r]$.

Table 1 Consonant phonemes of Iron (IPA and standard transliteration)

| | <i>Labial</i> | <i>Labiodental</i> | <i>Alveolar</i> | <i>Velar</i> | <i>Uvular</i> |
|-----------|----------------------------|----------------------|-------------------------------|--|--|
| Plosive | /p b p' / <i>p b p'</i> | | /t d t' / <i>t d t'</i> | /k k ^w g g ^w k' k' ^w / <i>k ku g gu k' k'u</i> | /q q ^w / <i>q qu</i> |
| Nasal | /m / <i>m</i> | | /n / <i>n</i> | | |
| Trill | | | /r / <i>r</i> | | |
| Fricative | | /f v / <i>f v</i> | /s z / <i>s z</i> | | /χ χ ^w ʁ ʁ ^w / <i>x xu γ γu</i> |
| Affricate | | | /ts dz tʃ / <i>c dz c'</i> | | |
| Lateral | | | /l / <i>l</i> | | |

‘farther’ (comparative). Initial *y*- before geminated *s* is not reflected orthographically: *ssædz* [ɬʃ:vɛz] ‘twenty’.

Vowels

The Ossetic vowel system can be divided into peripheral (strong) and central (weak) vowels (**Table 2**).

The vowels /u/ and /i/ have nonsyllabic variants that are rendered as *u* (sometimes *u*) and *j* in the transliteration. /u/ in onsets before vowels is realized as [w]. Epenthetic [w] is inserted between /u/ and any other vowel: *læu-* ‘stand’ and the infinitive marker *-yn* form *læuɰyn*. *j* is used as a glide between any vowel (except *u*) and *i/y*: *uda-* and *-yn* become *udaj-yn* ‘humidify’.

Accent

The word accent depends on the distribution of strong and weak vowels. If the first vowel is strong (s), it receives the accent, if it is weak (w), the second vowel is stressed. Thus, the following patterns emerge (accent is marked by an acute):

.śś .św .wś .wś

There are lexicalized exceptions to that rule (e.g., forms of the demonstrative pronoun and words like *Irón*). An emerging morphophonemic exception is the preverb *ys-* ([ɬʃ] or [ʃ]), which retracts the accent even with speakers who no longer articulate the initial vowel: (y)*s-æxgæn-yn* ‘to close’. Proper names are stressed on the second syllable, while retracting the accent to the initial syllable produces a pejorative note.

Retraction of the accent within a noun phrase (NP) marks the NP as definite (*zærdæ* ‘a heart’, *zærdæ* ‘the heart’).

Only scattered information is available about the phrasal accent of Ossetic. Abaev (1964) lists the noun

Table 2 Vowel phonemes of Iron (IPA and standard transliteration)

| | <i>Front</i> | <i>Central</i> | <i>Back</i> |
|------|-----------------|-----------------|-----------------|
| High | [i] <i>i</i> | [ɨ] <i>y</i> | [u] <i>u</i> |
| Mid | [e] <i>e</i> | | [o] <i>o</i> |
| Low | [a] <i>a</i> | [ɐ] <i>æ</i> | |

phrase (containing adjectives or genitives, *syrx tyrysa* ‘red flag’), postpositional phrases (*bælasɰ byn* ‘under the tree’), and complex predicates (*rox kænyn* ‘forget’) as phonological phrases. Enclitic pronouns and particles (such as negative *næ*) are also incorporated into phonological phrases.

Loan Word Phonology

The ejectives were apparently introduced through Caucasian loans (Iron *zač’æ*, Circassian [zatʃʰɛ] ‘beard’), although they also correspond to plain voiceless plosives in earlier Russian loans (Iron *bulk’on*, Russian *polkovnik* ‘colonel’). While older loans from Russian follow the Iron accent pattern, recent loans often preserve the lexical Russian accent. Also, Russian *s* [s] is sometimes realized as [ʃ] and sometimes as [s].

Nouns

Ossetic morphology is agglutinative with mildly inflectional elements. There are nine morphological cases which have, in part, developed from postnominal elements.

Subject and indefinite direct object are usually in the nominative (bare stem). Objects in the genitive are marked as definite. The dative marks the indirect

Table 3 Case system (for *kark* 'hen')

| | | <i>Singular</i> | <i>Translation</i> | <i>Plural</i> |
|-----------------------|----------|------------------|----------------------|---------------------|
| Grammatical cases | NOM | <i>kark</i> | 'hen(s)' | <i>kærc̣y-t-æ</i> |
| | GEN | <i>karč̣-y</i> | 'of the hen(s)' | <i>kærc̣y-t-y</i> |
| | DAT | <i>kark-æn</i> | 'to the hen(s)' | <i>kærc̣y-t-æn</i> |
| Local-adverbial cases | ALL | <i>kark-mæ</i> | 'to the hen(s)' | <i>kærc̣y-t-æm</i> |
| | ABL | <i>kark-æj</i> | 'from the hen(s)' | <i>kærc̣y-t-æj</i> |
| | SUPERESS | <i>kark-yl</i> | 'on the hen(s)' | <i>kærc̣y-t-yl</i> |
| | LOC | <i>karč̣-y</i> | 'at the hen(s)' | <i>kærc̣y-t-y</i> |
| Other adverbial cases | EQU | <i>kark-au</i> | 'as/than the hen(s)' | <i>kærc̣y-t-au</i> |
| | COM | <i>karč̣-imæ</i> | 'with the hen(s)' | <i>kærc̣y-t-imæ</i> |

object, but also the target or purpose of an action. The local cases express the primary local and temporal relations, but the ablative is also used to mark a tool or material used to perform an action, the superessive to mark a reason. The equative (EQU) marks the compared object with comparatives or the language in which something is written, said, etc. (*Iron-au* 'in Iron'), the comitative the partner involved in an action.

Plurals are formed by adding *-t-* to the stem plus the same case markers as in the singular. Sometimes, infixes are added after the stem, such as *-y-* in many cases where the stem ends in a consonant cluster (*cyxt* 'cheese', plural *cyxt-y-t-æ*) (Table 3).

Uninflected nouns function as adjectives, but there are also dedicated adjectives (*syddæg* 'clean'), sometimes marked by formatives like *-on* (*uarz-on* 'beloved' from *uarz-yn* 'love') or *-ag* (*xox-ag* 'mountainous' from *xox* 'mountain'). Adjectives and nouns used as adjectives take the comparative marker *-dær* (*dard-dær* 'farther') and stand in the superlative paraphrase with *æppæty* or *nuuyl* 'most' (*æppæty dard* 'farthest').

Pronouns

Pronouns inflect mostly like nouns. The personal pronouns have two stems, lack an inessive and a third person series, which is substituted from the remote demonstrative pronoun (Table 4).

The enclitic object pronouns lack a nominative and an equative to the effect that enclitically expressed direct objects have to be put in the genitive (Table 5).

The genitives of the full and enclitic personal pronoun and the reflexive pronoun substitute for the missing possessive pronouns. Reflexives are formed from the object pronoun with *-x-* and a set of special endings. For reciprocal expressions, the noun *kærædzi* 'one another', which only corresponds with plural antecedents, is used.

The demonstrative system exhibits a deictic split into remote (*u(y)-*) and local (*a-*). The true pronouns mark nominative and genitive by the same form

Table 4 Personal pronouns

| | <i>1 sg.</i> | <i>2 sg.</i> | <i>1 pl.</i> | <i>2 pl.</i> |
|----------|---------------|---------------|----------------|------------------|
| NOM | <i>æz</i> | <i>dy</i> | <i>max</i> | <i>symax</i> |
| GEN | <i>mæn</i> | <i>dæu</i> | <i>max</i> | <i>symax</i> |
| DAT | <i>mæn-æn</i> | <i>dæu-æn</i> | <i>max-æn</i> | <i>symax-æn</i> |
| ALL | <i>mæn-mæ</i> | <i>dæu-mæ</i> | <i>max-mæ</i> | <i>symax-mæ</i> |
| ABL | <i>mæn-æj</i> | <i>dæu-æj</i> | <i>max-æj</i> | <i>symax-æj</i> |
| SUPERESS | <i>mæn-yl</i> | <i>dæu-yl</i> | <i>max-yl</i> | <i>symax-yl</i> |
| EQU | <i>mæn-au</i> | <i>dæu-au</i> | <i>max-au</i> | <i>symax-au</i> |
| COM | <i>memæ</i> | <i>demæ</i> | <i>max-imæ</i> | <i>symax-imæ</i> |

(*a-j* 'this', *uy-j* 'that'), the other cases are formed by adding dative *-mæn* (*uy-mæn*), allative *-mæ*, ablative *-mæj* (*uy-mæj*), locative *-m*, superlative *-uyl*, equative *-jau*, and comitative *-imæ*. The plural forms *adon*, *uýdon* inflect like nouns. In adnominal position, an adjective is formed by adding *-cy* (*uýcy don* 'that water').

Interrogative pronouns inflect like the deictic pronouns and are split into personal (nominative *či* 'who', other cases *kæ-*) and impersonal (nominative *cy* 'what', other cases *cæ-*).

Numerals

The numeral system is basically a mixed decimal-vigesimal system, such that (1a) and (1b) are equivalent (Table 6).

(1a) *ærtyn fondz*
thirty five
 'thirty-five'

(1b) *fynddæ s æmæ (y)ssædz*
fifteen and twenty
 'thirty-five'

Ordinals are formed by means of a suffix *-æm* (*cyppar-æm* 'fourth'), distributives add *-gaj* (*iu-gaj* 'one by one').

Verbs

The Ossetic verb has a present stem and a past stem (ending in a dental stop). The former is the basis for

Table 5 Enclitic object pronouns

| | 1 sg. | 2 sg. | 3 sg. | 1 pl. | 2 pl. | 3 pl. |
|----------|-------------|-------------|---------------|-------------|-------------|----------------|
| GEN | <i>mæ</i> | <i>dæ</i> | <i>jæ, æj</i> | <i>næ</i> | <i>uæ</i> | <i>sæ</i> |
| DAT | <i>myn</i> | <i>dyn</i> | <i>(j)yn</i> | <i>nyn</i> | <i>uyn</i> | <i>syn</i> |
| ALL | <i>mæm</i> | <i>dæm</i> | <i>(j)æm</i> | <i>næm</i> | <i>uæm</i> | <i>sæm</i> |
| ABL | <i>mæ</i> | <i>dæ</i> | <i>dzy</i> | <i>næ</i> | <i>uæ</i> | <i>sæ, dzy</i> |
| LOC | <i>mæ</i> | <i>dæ</i> | <i>dzy</i> | <i>næ</i> | <i>uæ</i> | <i>sæ, dzy</i> |
| SUPERESS | <i>myl</i> | <i>dyl</i> | <i>(j)yl</i> | <i>nyl</i> | <i>uyl</i> | <i>syl</i> |
| COM | <i>memæ</i> | <i>demæ</i> | <i>jemæ</i> | <i>nemæ</i> | <i>uemæ</i> | <i>semæ</i> |

Table 6 Ossetic numerals

| Cardinal | Value |
|---------------|-------|
| <i>iu</i> | 1 |
| <i>dyuuæ</i> | 2 |
| <i>ærtæ</i> | 3 |
| <i>cyppar</i> | 4 |
| <i>fondz</i> | 5 |
| <i>æxsæz</i> | 6 |
| <i>avd</i> | 7 |
| <i>ast</i> | 8 |
| <i>farast</i> | 9 |
| <i>dæs</i> | 10 |
| <i>sædæ</i> | 100 |

Table 7 Major alternations in the past stem

| Present stem | Past stem |
|---------------------------|-----------|
| | -a- |
| -æ- | |
| -a- | -o- |
| -i-, -u-, -au-, -æu-, -o- | -æ- |
| -d-, -t-, -tt-, -nd-, -nt | -y- |
| -dz-, -c-, -ndz-, -nc | -s-t |
| -n-, -m | -γ-d |
| | -Ø-d |

the present and future tenses and all deverbal nouns, adjectives, and the infinitive (-*yn*). The latter forms the past tense and the past participle (bare stem).

The past stem shows facultative ablaut of the stem vowel and some facultative modifications of stem-final consonants, as in *lidz-* ‘run away’ (present) and *lyγ-d-* (past). -s- or -y- are sometimes inserted before the past stem marker (*zar-yn* ‘sing’, *zar-γd-t-æ-n* ‘I sang’). Transitive and intransitive verbs have different sets of past tense personal endings (Table 7).

The tense system distinguishes present (habitual, narrative, continuous present, and immediate future), past, and future.

In addition, the copula *uyn* distinguishes between a momentaneous (MOM) and a habitual (HAB) present. The third person present of the copula has the forms *u*, *i*, and *is*, which vary freely (Bagaev, 1965) (Table 8, Table 9).

(2) *Ucy don syγdæg u.*
that water clean be.3SG PRES MOM
 ‘That water is clean (right now).’

(3) *Ucy don syγdæg væjj-y.*
that water clean be-3SG PRES HAB
 ‘Such water is usually clean.’

Imperfective aspect is expressed lexically (*dzur-yn* ‘say’, *zæγ-yn* ‘tell’) or morphologically by adding one of the preverbs (generically *s-*). The preverbs also give a basic temporal-spatial orientation that takes into account the speaker’s position. They also express further notions of aspect and aktionsart (Table 10).

The subjunctive expresses doubt (present), wish, possibility (present and future), and necessity, and is used to give orders (future). The past subjunctive covers all these notions.

There are several constructions involving verbal nouns, such as the passive (past participle plus *cæu-yn* ‘go’) and the causative (infinitive plus *kæn-yn* ‘do’).

(4) *uarst cæu-y*
loved (past participle) go-PRES 3SG
 ‘she is loved’

Noun and Postposition Phrases

Nouns can be modified by means of a preceding noun in the genitive or an adjective. Many nouns can also function as adjectives:

(5a) *xur bon*
sun day
 ‘a sunny day’

(5b) *lædz-y cæsgom*
man-GEN face
 ‘the man’s face’

Coordinated elements show group inflection.

(6) *Æxsar æmæ Æxsærtædz-y rajguyrd*
Æxsar and Æxsærtæg-GEN birth
 ‘Æxsar’s and Æxsærtæg’s birth’

The postpositional constructions that express spatial and temporal relations usually involve functionally interpreted nouns (such as *sær* ‘head’

Table 8 Indicative verb forms (*kæn-/kod-* 'do' [tr.] and *kaf-/kafyd-* 'dance' [itr.])

| | <i>Present</i> | <i>Transitive past</i> | <i>Intransitive past</i> | <i>Future</i> |
|------|----------------|------------------------|--------------------------|----------------------|
| 1 SG | <i>kæn-yn</i> | <i>kod-t-on</i> | <i>kafyd-t-æn</i> | <i>kæn-dz-ynæn</i> |
| 2 SG | <i>kæn-ys</i> | <i>kod-t-aj</i> | <i>kafyd-t-æ</i> | <i>kæn-dz-ynæ</i> |
| 3 SG | <i>kæn-y</i> | <i>kod-t-a</i> | <i>kafyd-is</i> | <i>kæn-dz-æn</i> |
| 1 PL | <i>kæn-æm</i> | <i>kod-t-am</i> | <i>kafyd-yst-æm</i> | <i>kæn-dz-yst-æm</i> |
| 2 PL | <i>kæn-ut</i> | <i>kod-t-at</i> | <i>kafyd-yst-ut</i> | <i>kæn-dz-yst-ut</i> |
| 3 PL | <i>kæn-ync</i> | <i>kod-t-oj</i> | <i>kafyd-yst-y</i> | <i>kæn-dz-yst-y</i> |

Table 9 Subjunctive and imperative verb forms

| | <i>Subjunctive present</i> | <i>Subjunctive tr. past</i> | <i>Subjunctive intr. past</i> | <i>Subjunctive future</i> | <i>Imperative present</i> | <i>Imperative future</i> |
|------|----------------------------|-----------------------------|-------------------------------|---------------------------|---------------------------|--------------------------|
| 1 SG | <i>kæn-i-n</i> | <i>kod-t-a-i-n</i> | <i>kafyd-a-i-n</i> | <i>kæn-on</i> | | |
| 2 SG | <i>kæn-i-s</i> | <i>kod-t-a-i-s</i> | <i>kafyd-a-i-s</i> | <i>kæn-aj</i> | <i>kæn</i> | <i>kæn-iu</i> |
| 3 SG | <i>kæn-i-d</i> | <i>kod-t-a-i-d</i> | <i>kafyd-a-i-d</i> | <i>kæn-a</i> | <i>kæn-æd</i> | <i>kæn-æd-iu</i> |
| 1 PL | <i>kæn-i-kk-am</i> | <i>kod-t-a-i-kk-am</i> | <i>kafyd-a-i-kk-am</i> | <i>kæn-æm</i> | | |
| 2 PL | <i>kæn-i-kk-at</i> | <i>kod-t-a-i-kk-at</i> | <i>kafyd-a-i-kk-at</i> | <i>kæn-at</i> | <i>kæn-ut</i> | <i>kæn-ut-iu</i> |
| 3 PL | <i>kæn-i-kk-oj</i> | <i>kod-t-a-i-kk-oj</i> | <i>kafyd-a-i-kk-oj</i> | <i>kæn-oj</i> | <i>kæn-ænt</i> | <i>kæn-ænt-iu</i> |

Table 10 Directional preverbs

| | <i>Toward speaker</i> | <i>Away from speaker</i> |
|-----------------|-----------------------|--------------------------|
| Inward motion | <i>ærbə-</i> | <i>bə-</i> |
| Outward motion | <i>ra-</i> | <i>a-</i> |
| Downward motion | <i>æf-</i> | <i>nyn-</i> |

for 'top') that additionally take one of the case endings. The dependent noun then receives the genitive marker.

- (7) *xox-y* *sær-yl*
mountain-GEN *head-SUP*
 'on top of the mountain'

A construction with an adnominal genitive noun can be paraphrased as dative with a clitic pronoun in the genitive.

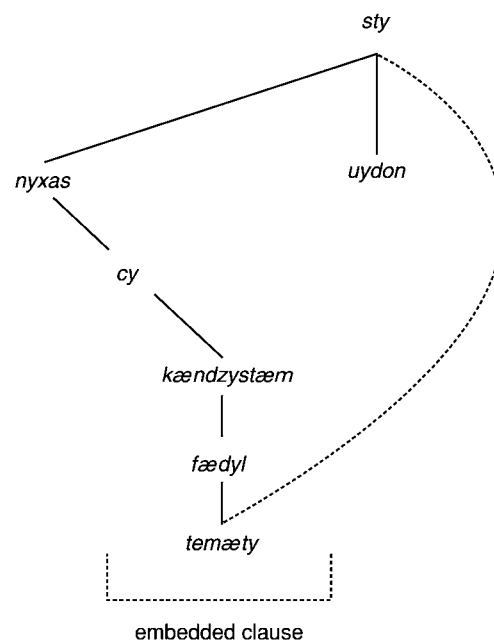
- (8a) *Nart-y* *fyr*
Nart-GEN *son*
 'son of the nart'
 (8b) *Nart-æn* *jæ* *fyr*
Nart-DAT *he.GEN* *son*
 'son of the Nart'

Simple Verbal Sentences

In most cases, the arguments precede the verb (SOV order).

- (9) *Nart* *udævdz* *fyng-yl* *sæværd-t-oj*.
Nart *shawm* *table-SUP* *put-PAST-3PL*
 'The Nart put the shawm on the table.'

In focused word order, the verb can precede the subject. There are no expletive subjects, thus the most simple type of a verbal sentence contains just a verb.

**Figure 2** Dependency markers of sentence (14).

- (10) *uar-y*
rain-PRES 3SG
 'it is raining'

Since subjects can be dropped, intransitive verbs can also form one-word sentences.

- (11) *xau-y*
fall-PRES 3SG
 'he/she/it falls (is falling)'

Clitic objects (always attached to the first phrase of a sentence) stand in for an omitted object or an adverbial noun (12a), or they are presumptive (12b).

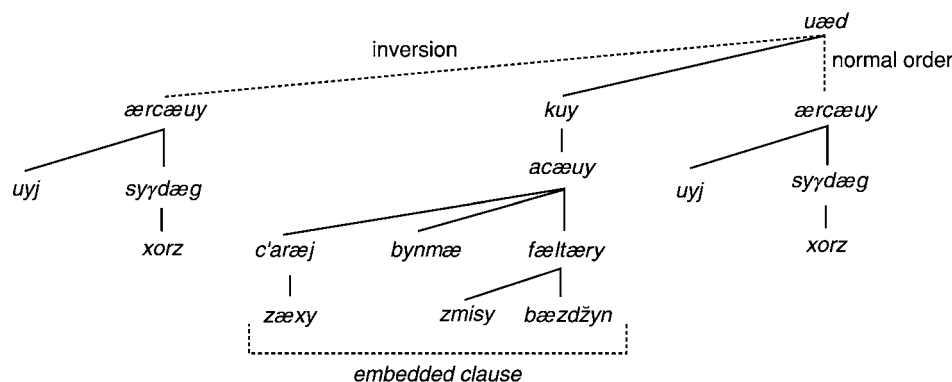


Figure 3 Dependency markers of sentence (15).

- (12a) Nart yl udævdz sæværd-t-oj.
 Nart it.SUP shawm put-PAST-3PL
 'The Nart put the shawm on it.'
- (12b) Nart yl udævdz
 Nart it.SUP shawm
 sæværd-t-oj fyng-yl.
 put-PAST-3PL table-SUP
 'The Nart put the shawm on it, on the table.'

The *constructio ad sensum* is very common for both singular subjects with plural verbs and vice versa.

Copular Sentences

Sentences with the copula *uyn* have the word order (a) subject, predicate noun, copula or (b) subject, copula, predicative noun.

- (13) Mæ nom u Zæhra.
 my name be.3SG PRES MOM Zæhra
 'My name is Zæhra.'

The copula can combine with preverbs: *s-uyn* 'become' and *fæ-uyn* 'turn out to be'.

Syntax of Embedding

We give two sample analyses of embedding constructions. Example (14) (Figure 2) shows a relative clause with a pseudo-antecedent (agreeing in number with the main verb) nested inside the relative clause. Example (15) (Figure 3) illustrates a common construction with attributive clauses and conditionals. Such clauses usually precede the main clause. If the order is inverted, the correlative word (pronoun or conjunction) is moved to the very end of the sentence

behind the dependent clause (main clauses in bold print):

- (14) **Nyxas** cy temæ-t-y
 talk what subject-PL-GEN
 fædyl kæn-dzyst-æm, **uydon** st-y.
 about do-FUT-1 PL those be-3 SG
 ('The talk about which subjects we are going to make are these.')
- The subjects about which we are going to talk are these.
- (15) **Uyj** xorz syrdæg ærcæu-y,
 that good clean arrive-3SG
 zæxx-y c'ar-æj byn-mæ
 earth-GEN crust-ABL ground-ALL
 zmis-y bæzdžyn fæltær-y
 sand-GEN thick layer-INESS
 kuy acæu-y, uæd.
 when come out-3SG then
 'When it (the water) comes out from the earth's crust to the ground through a thick layer of sand, then it arrives fairly clean.'

Bibliography

- Abaev V I (1964). *A grammatical sketch of Ossetic*. The Hague: Mouton.
- Bagaev N K (1965). *Sovremennyy ossetinskij iazyk*. Vladikavkaz: Severo-osetinskoe knizhnoe izdatel'stvo.
- Miller V I (1927). *Osetinsko-russko-nemetskij slovar'*. St. Petersburg: Akademiia Nauk SSSR.
- Testen D (1997). 'Ossetic phonology.' In Kaye A S (ed.) *Phonologies of Asia and Africa (including the Caucasus)*. Winona Lake, Ind.: Eisenbrauns. 707–731.
- Thordarson F (1989). 'Ossetic.' In Schmidt R (ed.) *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert. 456–479.

Oto-Mangean Languages

T Kaufman, University of Pittsburgh, Forest Hills, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Oto-Mangean (OM) is the most temporally diverse genetic grouping of languages spoken within Meso-America and one of the most widespread geographically. Currently, Mayans account for more speakers and more territory, but the individual identities of the approximately 30 distinct languages of Mayan are not in doubt, whereas the exact number of languages under the names Otomí, Chinanteko, Popoloka, Masateko, Sapoteko, Chatino, and Misteko is a matter of continued discussion and debate. The number recognized here (ca. 30) is probably close to a minimum number.

Oto-Mangean is a stock of roughly the time depth of Indo-European – approximately 6000 years. It is made up of seven readily recognizable language families, some of them with fairly complicated ramification and individually of varying time depth: Oto-Pamean (3600 years), Chinanteko (1500 years), Chorotegan (1300 years), Tlapanekan (800 years), Masatekan (2500 years), Sapotekan (2400 years), and Mistekan (3700 yrs). There is one language, Amusgo, that does not form part of a family, but the closest relatives of Amusgo are the languages of the Mistekan family. The previous groupings are generally agreed on.

The internal makeup of OM has been known since approximately the turn of the 20th century. Connections across these families and isolates have been observed since late in the 19th century, but the existence of Oto-Mangean as we now know it (roughly) was outlined basically during the 1920s.

Intermediate groupings are still being worked out. Comparative phonology and especially comparative grammar studies done by the author (Kaufman, 1983, 1988) show that there are two levels of ramification between the individual families and the ancestral proto-Oto-Mangean (pOM): The major splits are called ‘divisions’ and the groupings under the divisions are ‘branches.’ OM has two divisions, eastern and western. The eastern and western divisions have two main branches each: The eastern division (4700 years) contains Masatekan-Sapotekan (3500 years) and Amusgo-Mistekan (?4000 years), and the western division (4700 years) contains Oto-Pamean-Chinanteko (4000 years) and Chorotegan-Tlapanekan (4000 years). Each division is as diverse as Yuta-Nawan (Uto-Aztecan), and each branch is as diverse as Mayan; some of the families within OM (Mistekan

and Oto-Pamean) are more diverse than Mije-Sokean (Mixe-Zoquean). The various families of OM are like the language groups of Indo-European, such as Indic, Iranian, Baltic, Slavonic, Germanic, Celtic, and Romance. Regarding one of the families, Mistekan, there has been some question about whether to include or exclude Amusgo and Triki.

Morris Swadesh mounted an argument that Wavi belongs to the OM stock; this hypothesis was cautiously accepted by Longacre, but no later Oto-Mangeanist has found the hypothesis valid or even promising.

The following classification gives:

- Name of language or genetic group: Okwilteko, Oto-Mangean
- Favored Spanish orthography: <Ocuilteco, Otomangue>
- Synonyms: SYN
- Lexicostatistic time depth calculated by Kaufman, etc. (mc = minimum centuries) [NNmc Kaufman]
- Number of speakers reported in 1990 Mexico census: Mx
- Number of speakers reported in Ethnologue 2002: EL
- Country or state where spoken: COSTA RICA, GUANAJUATO.

In the following classification, in which language areas are named and not further subdivided, many researchers will recognize two or more distinct (emergent or virtual) languages. This is especially true in the cases of ‘Misteko’ and ‘Sapoteko.’

- Oto-Mangean stock (Sp. <Otomangue>) [60mc Kaufman] MEXICO, NICARAGUA, COSTA RICA
- Western Oto-Mangean division [47mc Kaufman]
- Oto-Pamean-Chinanteko branch [40mc Kaufman]
- Oto-Pamean family [36c Kaufman]
- Pamean (northern Oto-Pamean) subfamily [25c Kaufman]
- Chichimeko language (Sp. <Chichimeco>, <Jonaz>) EL: 200 GUANAJUATO
- Pame Complex [14c Kaufman] Mx: 5.6k SAN LUIS POTOSI
 1. Northern Pame virtual language EL: 1–10k
 2. Central Pame virtual language EL: 4k
 3. Southern Pame virtual language
- Southern Oto-Pamean subfamily [24c Kaufman]
- Matlatzinka-Okwilteko language area [8–9c Kaufman] STATE OF MEXICO
- Matlatzinka emergent language (Sp. <Matlatzinca>) EL: ca. 30

- Okwilteko emergent language (Sp. <Ocuilteco>, <Tlahuica>) EL: 50–100
- Otomíán group [10c Kaufman]
- Otomí language area [8c Kaufman] Mx: 306k; EL: 223k
 1. Northeast Otomí emergent language VERA-CRUZ
 2. Northwest Otomí emergent language HIDALGO
 3. Western Otomí emergent language QUERETARO, MICHOACAN, Colonial
 4. Tilapa Otomí emergent language STATE OF MEXICO
 5. Ixtenco Otomí emergent language TLAXCALA
 6. Jalisco Otomí [extinct: undocumented] JALISCO
- Masawa language (Sp. <Mazahua>, COL Mazateco) Mx: 194k; EL: 365–370k
- Chinanteko family (Sp. <Chinanteco>) [15mc Swadesh] OAXACA Mx: 77.1; EL: 86.7–87.7k
 1. Ojitlán (N) Chinanteko language
 2. Usila (NW) Chinanteko language
 3. Quiotepec (W) Chinanteko language
 4. Palantla (LL) Chinanteko language
 5. Lalana (SE) Chinanteko language
 6. Chiltepec (EC) Chinanteko language
- Tlapaneko-Mangean branch (Sp. <Tlapaneco-Mangue>) [40mc Kaufman]
- Tlapaneko-Sutiaba language area (Sp. <Tlapaneco-Subtiaba>) [8mc Swadesh] Mx: 55.1k; EL: 66.7k
 1. Malinaltepec (general) Tlapaneko emergent language (SYN Yopi) GUERRERO
 2. Azoyú (orig. <Atzoyoc>) Tlapaneko emergent language GUERRERO
 3. Sutiaba (orig. <Xoteapan>) emergent language
 4. (Sp. <Subtiaba>, <Maribio>) NICARAGUA
- Chorotegan family (Sp. <Mangueano>, <Chiapaneco-Mangue>) [13mc Swadesh]
 1. Chiapaneko language (Sp. <Chiapaneco>) CHIAPAS
 2. Chorotega (Mange) language (Sp. <Mangue>, <Orotiña>, <Chorotega>, <Choluteca>) HONDURAS, NICARAGUA
- Eastern Oto-Mangean division (Sp. <Otomangue oriental>) [47mc Kaufman]
- Masatekan-Sapotekan branch [35mc Kaufman]
- Masatekan family (Sp. <Polopocano>) [24c Swadesh]
- Masateko complex (Sp. <Mazateco>) [10c Swadesh] Mx: 124.2k; EL: 174.5k
 1. Huautla-Mazatlán Masateko language EL: 50–60k OAXACA
 2. Ayautla-Soyaltepec Masateko language EL: 40k OAXACA, PUEBLA
 3. Jalapa Masateko language EL: 10–15k OAXACA, VERACRUZ
 4. Chiquihuitlán Masateko language EL: 3–4k OAXACA
- Chochoan subfamily [12c Swadesh]
- Iskateko language (Sp. <Ixcateco>) EL: <50 OAXACA
- Chocho-Popoloka language area (Sp. <Popoloca>) 8c
 1. Chocho emergent language Mx: 12.3k; EL: 428 OAXACA
 2. Northern Popoloka emergent language PUEBLA
 3. Western Popoloka emergent language PUEBLA
 4. Eastern Popoloka emergent language PUEBLA
- [Popoloka: Mx: 23.8k; EL: 23.2k]
- Sapotekan family (Sp. <Zapotecano>) [24c Swadesh]
- Sapoteko complex (Sp. <Zapoteco>) [14c Rendón] Mx: 423k; EL: 326k OAXACA
 1. Northern Sapoteko language area
 2. Central Sapoteko language area
 3. Southern Sapoteko language area
 4. Papabuco Sapoteko language area
 5. Western Sapoteko language area
- Chatino language area Mx: 20.5k; EL: 36k OAXACA
 1. Yaitepec Chatino emergent language
 2. Tatalpepec Chatino emergent language
 3. Zenzontepec Chatino emergent language
- Amuzgo-Mistekan branch (Sp. <Amuzgo-Mixtecano>)
- Amuzgo language (Sp. <Amuzgo>) Mx: 1.7k; EL: 28.2k OAXACA, GUERRERO
- Mistekan family (Sp. <Mixtecano>) [37c Kaufman]
- Misteko-Kwikateko subfamily (Sp. <Mixteco-Cuicateco>) [25c Swadesh]
- Misteko complex (Sp. <Mixteco>) [15c Swadesh] Mx: 323.1k; EL: 327k
 1. Northern Misteko language area OAXACA, PUEBLA
 2. Central Misteko language area OAXACA
 3. Southern Misteko language area OAXACA, GUERRERO
- Kwikateko language (Sp. <Cuicateco>) Mx: 14.2k; EL: 18.5k OAXACA

- Triki language area (Sp. <Trique>, <Triqui>) [10mc Kaufman] Mx: 8.4k; EL: 23k OAXACA
 1. Chicahuaxtla Triki emergent language
 2. Copala Triki emergent language
 3. Itunyoso Triki emergent language

The homeland of the OM languages seems to have been somewhere in the highland part of Meso-America between western Oaxaca and the basin of Mexico. The OM languages began to break up after the early stages of domestication of some plants approximately 7000 years ago and long before the transition to agriculture approximately 2000 B.C.E. The extent of the homeland is difficult to gauge because, in principle, if it is large it is going to have internal diversification, especially in broken country such as the highlands of Oaxaca and Puebla. Part of Chinanteco country (the Chinantla) in Oaxaca is particularly lush and fertile and would have been favored by early populations who did not have much competition from rival groups. Before the transition to agriculture, most Meso-American populations were small and much territory was unoccupied.

Structural Characteristics of Oto-Mangean Languages

Phonological Traits

OM languages vary quite widely in the number of contrastive vowels and the complexity of syllable onsets. In the number of contrastive consonants, tones, and syllable codas, the languages differ less. We first characterize for related sets of traits how pOM phonology was developed (as outlined in Kaufman, 1988) and then discuss some of the ways that the lower-level protolanguages and individual languages deviate from the original patterns.

Consonants

pOM has five plosives /p t tz k kw/, four spirants /s j jw/, two laryngeals /ʔ h/, one lateral /l/, two nasals /m n/, and two semivowels /y w/. <tz> (or <c>) is a sibilant affricate [ts]. <8> is 'theta'; an equally plausible phonetic reconstruction is [r]. <j> (or <x>) is [x], and <jw> (or <xw>) is [xw]. There is little evidence for reconstructing *p, but it seems required for the protolanguage. Most Meso-American languages have /p/ and lack /kw/, and labialized velars generally. pOM *kw has shifted to [p] in several languages, as discussed later. The evidence for *m is better than for *p, but many OM languages lack /m/ as a phoneme.

Vowels and Syllabic Nuclei

pOM has four vowels /i e a u/ and five complex nuclei /ia ea ua ai au/. /ia/ (or <4>) may be 'barred,' /ea/ (or <6>) may be 'schwa,' /ua/ may be [o], /ai/ (or <3>) may be 'aesc,' and /au/ (or <2>) may be 'open o.' Among extant OM languages, the smallest system has four vowels /i e a o/ (some forms of Zapotec) and the most elaborate has nine vowels /i e 3 4 6 a u o 2/ (some forms of Otomi). pOM syllables can close with nothing, /n/, /ʔ/ (glottal stop), /h/, /nh/, /nʔ/, or /nhʔ/. Syllable-closing *n is realized as vowel nasality where it survives; some languages, discussed later, have lost nasality on vowels. Not all languages have a clear reflex of syllable-closing *h. pOM probably had three level tones and possibly a rising and a falling tone. Chorotegan seems to lack tone, but the majority of the remaining languages have a pattern analogous to the suggested reconstruction. A few have four level tones along with the moving tones; several have only two or three tonal contrasts altogether.

Syllable Onsets

These are fairly complex for pOM. A complex onset is set up as a plosive or resonant preceded by *n, *y, *h, *ʔ, or a combination of these (although of resonants, apparently only *l can be preceded by *n). (T = plosive /p t tz k kw/, R = resonant /m n l/, H = laryngeal /ʔ h/.) They are set up this way even though in some branches and individual languages the reflexes of *h *ʔ appear following the plosive; *y always appears following the plosive or resonant and *n always precedes. The maximal onset is *nyHC, which is [nTyH] where C is a plosive and [HRy] where C is a resonant. The reason for this analysis is that all these preposed consonants appear as the exponents of prefixed, mostly derivational, morphemes, although any particular instance of one of these preposed consonants may not be segmentable. Since there is no contrast between /yT/ and /Ty/, /ʔT/ and /Tʔ/, etc., this analysis is maximally general and consistent. Some languages (Zapotecan and Mixtecan) have eliminated *h or *ʔ in complex onsets, some have eliminated only *ʔ (Tlapanecanand Amusgo), some have eliminated *n in complex onsets (Zapotecan), some realize *ʔT = /Tʔ/ through (allophonic) glottalization, and some realize *hT = /Th/ through (allophonic) aspiration. In some languages, *yC has yielded palatalized consonants or their further developments.

Morpheme patterns are V, CV, VCV, CVCV, where C is any of the permitted syllable onsets, and V is any of the permitted vocalic nuclei and 'open' syllables.

Certain OM families (OP and Chin) underwent phonological change that rendered most stems

monosyllabic; the remaining families have stems that are (or were) predominantly disyllabic (Masatekan, Triki, Amusgo, and Southern Chatino change many CVCV stems to CCV, especially when V1 is a high vowel); most disyllabic stems, however, are morphologically complex, consisting of a monosyllabic root preceded by another root or a derivational prefix. This means that there is a much greater than even chance that when disyllabic stems of similar or identical meaning are compared across OM branches, if they are cognate they will be cognate for the first or last morpheme only. This, along with the great time depth within the stock, makes finding valid etymologies quite a challenging undertaking. A thorough understanding of word formation processes for each language, including the now no longer productive ones peculiar to each of the families, is needed before effective cognate searches can be made.

Quite a few comparative studies designed to identify cognates and reconstruct ancestral phonology and semantics have been carried out within OM. Most have been flawed by a failure to recognize many or most of the now unused word formation processes of earlier stages.

Grammatical Traits of Oto-Mangean Languages

This discussion is presented in terms of what is most prevalent and at the same time most probably the earliest state of OM languages; less common or more recent patterns are discussed later or not at all.

OM languages are consistently left-headed, with VO word order; any affixes are prefixes. Most grammatical morphemes are clitics; some are phonologically full words.

If we take ‘derivation’ to include both the formation of new lexical items and the change of morphosyntactic class without creating new lexical items, the following patterns are observable: prefixes can mark active, nonactive, and mediopassive ‘voice,’ versive (‘inchoative’) intransitive verbs based on nouns and adjectives, and causative verbs based on intransitive (including versive) and transitive verbs. All of the previous functions, except possibly ‘active,’ can be marked with grammatical words (often clitics) in one or another language. Left-headed compounds of the form NN and NA are found in most languages, and verbs of the shape VN with incorporated nouns are widely attested.

Much cliticization can be mistaken for inflexion, and in some cases inflexion versus cliticization cannot be decisively determined. However, suffixation and unproblematic inflexion are found only in Oto-Pamean.

Pronouns are noun phrase (NP) substitutes: in most languages, when a noun argument is present, no pronominal will be present. Most pronouns have both a full and a cliticized phonological form. Nevertheless, only a few languages have redeployed cliticized third-person pronouns as agreement markers that may occur in the same clause with fully NP arguments. Non-third-person pronouns may mark an exclusive:inclusive distinction in the first person, a humble versus prideful distinction for the first person, and a polite versus familiar distinction in the second person.

Predominant constituent order includes verb–subject–object (VSO), prepositions (Pr), noun–genitive (NG), noun–adjective (NA), noun–demonstrative (ND), Quantifier–noun (QN), and noun–relative clause (NR).

Alignment

There are apparently three types of NP role alignment in OM languages, marked by distinct sets of pronominal markers for each but the last type: ‘ergative’ (e.g., Chinantec and probably Tlapanec), ‘active’ (e.g., Chocho, Matlatzincan, and probably Chiapanec), and ‘undifferentiated’ (Mixtec and Zapotec) (only one set of pronominal markers). No clear instances of accusative alignment have been identified.

Verbs mark aspect and mood by means of preposed morphemes that are clitics by origin, although in some languages they are indistinguishable from prefixes. Tense (time) is not marked, or it is marked by adverbs that are not positioned as are aspect and mood markers. Virtually all languages have a verb form (‘dependent’) that is used when the verb is subordinated to a higher predicate. The dependent form may also be used in dependent clauses expressing the function of optative, possible future, or, in some cases, imperative.

Locative adpositionals are encoded by means of body-part nouns and other nouns that denote the parts of things: thus, in/inside = belly, on [surface] = face/eye, on [top] = head, under = butt, between = interval, etc. This pattern is widespread although not universal in Meso-American languages. Other semantic connexions include tip = nose, leg [of table or chair] = foot, front = face, and edge = lip/mouth. As stated previously, adpositionals in OM languages are preposed.

Possession Classes of Nouns

Many OM languages subdivide nouns by the ways they mark types of possession; for example, a fair number of nouns will be accompanied by one or another grammatical morpheme that correlates with the fact that the noun is possessed.

Endocentric Noun Classes

Some OM languages, such as Zapotecan, Mazatecan, and Mixtecan, have a series of noun classifiers that mark such categories as 'tree,' 'fruit,' and 'animal.' They are proclitics; in some cases, their origin as independent nouns can be discerned, in others it cannot, but they do not qualify as prefixes. These classifiers occur before the noun that they classify and are therefore not like (other noun) modifiers. They may in fact be the heads of the constructions in which they occur. In some languages, some of their uses are optional. In all languages, some of their uses are lexicalized – the lexeme does not occur without the classifier, or without the classifier the noun has a different meaning than it does with the classifier. This pattern is possibly an old one dating back to the eastern OM level since some of the classifying morphemes are cognate and do not exist as independent lexical items (although some are not cognate across families or can be related to independent lexical items: These would show the effects of analogy and renewal).

Exocentric Noun Classes in Oaxacan Languages

OM languages are spoken in several zones, among which are the northern fringe of Meso-America, the basin of Mexico, southern Guerrero, the Tehuacan Valley, the Mixteca Alta, and Oaxaca, of which the last two are the most momentous linguistically. OM languages of Oaxaca belonging to several branches, including all Zapotecan, some Mixtecan, and some Mazatecan languages, assign nouns to several classes, which are marked by the third-person pronouns that refer to the nouns of the various classes. The classes that are found are semantically motivated and include such categories as 'adult human,' 'man,' 'woman,' 'irrational (baby, foreigner),' 'animal,' 'thing,' and 'god.' In virtually all cases, the pronouns (they are not agreement markers) used to mark these classes are phonologically reduced/simplified forms of nouns naming the corresponding semantic field. This phenomenon therefore seems to have originated within the past 1000–1500 years in a continuous region and to not be the continuation of a pattern present in any of the family-level proto-languages.

Precolumbian Language Contact with Non-Oto-Mangean Languages

Yokel Anxiety

Although the earliest surviving organic material pointing to (incipient) maize domestication (dating

to approximately 7000 years ago) has been found in dry caves in the state of Puebla, firmly within the OM area, maize domestication was developing as well in lowland areas where the organic material simply did not survive. At any rate, Mayan speakers (in the Mayan lowlands), Mixe-Zoquean speakers (in Olmec country and in the basin of Mexico), and Totonacan speakers (in the basin of Mexico) developed complex society slightly before Zapotecan speakers (in the valley of Oaxaca), Chorotegan speakers (near Cholula), Matlatzincan speakers (near Toluca), or Mixtecan speakers (northwestern Oaxaca). Most of the OM languages spoken by long-term practitioners of complex society in Meso-America show serious phonological and lexical influence from such non-Otomangean languages as Mixe-Zoquean and Mayan. Unlike the oldest state of OM languages, Mije-Sokean, Totonakan, and Mayan languages lack /kw/, vowel nasality, and tone, and they generally have predictable stress. In imitation of non-OM languages, whose speakers had higher prestige, the following changes were adopted by OM languages through what I call 'yokel anxiety': (1) Oto-Pamean and Zapotec (but not Chatino) changed *kw to [p]; (2) Matlatzincan, Chorotegan, and Zapotec (but not Chatino) eliminated nasality from vowels; (3) Chorotegan eliminated tone (or reduced it to a two-way stress contrast); and (4) Zapotec (but not Chatino) and Mixtec-Cuicatec imposed a penult-syllable stress pattern on its inherited and surviving tone systems of three or four tones.

The Mayanization of Oto-Pamean

The homeland of Oto-Pamean was the basin of Mexico. Between 2000 and 1500 B.C.E., it was bordered to the east by the Mayan language that developed into the ancestor of Huastec and Cabil (Chicomuselte). At that time, Oto-Pamean shifted its inherited *kw to [p], in imitation of Mayan, which has [p] and lacks [kw], and borrowed a few Mayan lexical items. By 1500 B.C.E., Oto-Pamean broke up into northern Oto-Pamean (or Pamean) and southern Oto-Pamean (or Otomian). Pamean expanded northward into the western parts of the state of San Luis Potosí and the southern part of the state of Zacatecas. Pamean came into contact with undocumented (and now extinct) languages beyond the northern border of Meso-America. Approximately 1000 C.E., the Pamean-speaking area dried out through climate change and agriculture was no longer possible. Pameans became foragers and occasional raiders on their Meso-American cousins to the immediate south, and they adopted some linguistic traits from their non-Meso-American northern neighbors. The (Otomian-speaking part of) basin of Mexico

came under serious Mayan (probably Tabasco Chontal or Yokot'an) influence in the Epiclassic period (ca. 700–1000 C.E.). Besides some lexical influence, the grammatical effect on Otomian was extensive: VOS word order, AN word order, and marking of the person of actors and possessors by preposed morphemes were all modeled on Mayan grammatical patterns. Since Pamean shares the last trait, and it is probable that the SOV, GN, and NA orders, as well as the presence of postpositions in Pamean are an adjustment to non-Meso-American (perhaps Hokan) languages, it is possible that Pamean earlier shared the Mayanized VOS and AN orders with Otomian. If so, this pattern would have been due to contact between proto-Oto-Pamean and pre-Huastecan.

Viability

In pre-Columbian times, most OM populations (except northern Oto-Pameans = Pameans) were agricultural, communities were sedentary, and the total population for each language was at least 10 000, increasing to more than 500 000. At present,

- Four OM languages have died out: South Pame, Sutiaba, Chiapanec, and Chorotega. Jalisco Otomi was never documented.
- Many OM languages are dying (moribund) (spoken only by elderly people); for example, Ocuiltec, Chocho, and Ixcatec.
- Most other OM languages are dwindling (obsolescent) (not being learned by children); for example, Matlatzinca and Popoloca.

A number of OM languages are merely endangered (being learned by children (with some attrition) with strong pressure for bilingualism in Spanish and pressure to abandon the native language), including some varieties each of Otomi, Chinantec, Tlapanec, Mazatec, Mixtec, Zapotec, and Chatino. Otomi (306 → 223k), Mazahua (194 → 363k), Mixtec (322 → 327k), and Zapotec (423 → 326k) (Mexican census 1990 → Ethnologue 2002) each have more than 225 000 speakers. If current trends continue, in 100 years probably only some varieties of these four will still be spoken in some fashion.

Documentation

Many OM languages were documented during the colonial period (1519–1814) by Catholic missionaries who wanted their replacements to have the ability to communicate Christian teaching to the Indian population, who had mostly been forcibly converted. This documentation included grammars and dictionaries of often more than 5000 entries,

along with translations of the catechism, confessional, and sermons and narratives from the Bible. The orthography was often inadequate to express all sounds accurately, and the grammatical models were often simply calqued on the traditional analysis of Latin. Especially valuable documentation for Misteko, Sapoteko, Matlatzinka, and Otomi has survived to the present day. Since 1930, Protestant missionaries, mostly English-speaking from the United States, Canada, and Great Britain and belonging to the Summer Institute of Linguistics/Wycliffe Bible Translators, have been working on Meso-American Indian languages, and most OM languages have been documented rather fully by them, especially by dictionaries and in many cases by grammars, less so by texts. Since 1950, a number of academic linguists, both Mexicans and English speakers, have worked on the documentation of OM languages. Much has been accomplished, but much remains to be done, especially with regard to dying and dwindling languages.

Bibliography

- Bartholomew D A (1965). 'The reconstruction of Otopamean (Mexico).' Ph.D. diss., University of Chicago.
- Benton J (1988). 'Proto-Zapotec phonology.' M.A. thesis.
- Bradley C H & Josserand J K (1978). *El Protomixteco y sus Decendientes*. Unpublished manuscript.
- Fernández de Miranda M T (1951). 'Reconstrucción del Protopoloca.' *Revista Mexicana de Estudios Antropológicos* 12, 61–93.
- Fernández de Miranda M T (1995). *El Protozapoteco*. Mexico DF: COLMEX and INAH.
- Fernández de Miranda M T & Weitlaner R J (1961). 'Sobre algunas relaciones de la familia Mangue.' *Anthropological Linguistics* 3(1).
- Gudschinsky S (1959). *Proto-Popotecan, a comparative study of Popolocan and Mixtecan*, Indiana University Publications in Anthropology and Linguistics No. 5. Baltimore: Waverly.
- Hopkins N A (1984). 'Otomanguean linguistic prehistory.' In Josserand J K, Winter M & Hopkins N A (eds.) *Essays in Otomanguean cultural history*, Vanderbilt University Publications in Anthropology No. 31. Nashville, TN: Vanderbilt University. 25–64.
- Josserand J K (1983). 'Mixtec dialect history.' Ph.D. diss., Tulane University.
- Kaufman T (1983). New perspectives on comparative OtoManguean phonology. Unpublished manuscript.
- Kaufman T (1988). OtoManguean tense/aspect/mood, voice, and nominalization markers. Unpublished manuscript.
- Kaufman T (1990a). Early OtoManguean homelands and cultures: some premature hypotheses. *University of Pittsburgh Working Papers in Linguistics* 1, 91–136.
- Kaufman T (1990b). Tlapaneko-Sutiaba, OtoManguean, and Hokan: where Greenberg went wrong. Unpublished manuscript.

- Kaufman T (1993). 'The native languages of Latin America.' In Moseley C & Ashley R E (eds.) *Atlas of the world's languages*. London: Routledge. 31–76.
- Kirk P L (1966). 'Proto Mazatec phonology.' Ph.D. diss., University of Washington.
- Longacre R E (1957). *Proto-Mixtecan*, Indiana University Research Center in Anthropology, Folklore, and Linguistics No. 5. Bloomington: Indiana University Press.
- Longacre R E (1961). 'Swadesh's Macro-Mixtecan hypothesis.' *International Journal of American Linguistics* 27, 9–29.
- Longacre R E (1962). 'Amplification of Gudschinsky's Proto-Popolocan-Mixtecan.' *International Journal of American Linguistics* 28, 227–242.
- Longacre R E (1966a). 'On linguistic affinities of Amuzgo.' *International Journal of American Linguistics* 32, 46–49.
- Longacre R E (1966b). 'The linguistic affinities of Amuzgo.' *Summa Anthropologica en homenaje a Roberto J. Weitlaner*. 541–560.
- Longacre R E (1967). 'Systemic comparison and reconstruction.' *HMAI* 5, 117–135, 141–142, 145–159.
- Mak C & Longacre R (1960). 'Proto-Mixtec phonology.' *International Journal of American Linguistics* 26, 23–40.
- Rensch C R (1963). 'Proto-Chinantec phonology.' M.A. thesis, University of Pennsylvania.
- Rensch C R (1966). 'Comparative Otomanguean phonology.' Ph.D. diss., University of Pennsylvania. *Indiana University Language Science Monographs* 14 (Published in 1976).
- Rensch C R (1973). 'Otomanguean isoglosses.' *Current Trends in Linguistics* 11, 295–316.
- Rensch C R (1977). 'Classification of the Otomanguean languages and the position of Tlapanec.' In *Two studies in Middle American comparative linguistics*. Arlington, TX: Summer Institute of Linguistics. 53–108.
- Rensch C R (1989). *Studies in Chinantec Languages* No. 1. *An etymological dictionary of the Chinantec languages*, Arlington, TX: Summer Institute of Linguistics.
- Swadesh M (1947). 'The phonemic structure of Proto-Zapotec.' *International Journal of American Linguistics*.
- Swadesh M (1960). 'The Oto-Chorotegan hypothesis and macro-Mixtecan.' *International Journal of American Linguistics* 26, 79–111.
- Swadesh M (1964a). 'Interim notes on Oaxacan phonology.' *Southwestern Journal of Anthropology* 20, 169–189.
- Swadesh M (1964b). 'Algunos problemas de la lingüística otomangue.' *Anales de Antropología* 1, 91–123.

This page intentionally left blank

P

Pahlavi

A V Williams, University of Manchester,
Manchester, UK

© 1994 Elsevier Ltd. All rights reserved.

Pahlavi is generally synonymous with the term 'Middle Persian,' i.e., the language of the Sasanians (224–651 AD) and their subjects in the province of Fars in southwest Iran. It was imposed by the Sasanian authorities as the sole official language of Iran and consequently it became the living language of the state religion, Zoroastrianism. More specifically, 'Pahlavi' (Book Pahlavi) is the name given to the medieval language in which Zoroastrian religious texts were written down in the Sasanian and early Islamic periods until the tenth century AD. It may be compared with the other main Middle Iranian dialects: Parthian, Soghdian, Khwarazmian, and Khotanese Saka. The earliest examples of Pahlavi script are found in rock inscriptions and on coins from the Parthian and early Sasanian periods. The orthographical system was derived from the western semitic consonantal script of Aramaic, the court language of the Achaemenian Empire. The cursive script of Book Pahlavi is full of ambiguities and corrupt forms, having only 14 letters (as compared with the 22 letters of Imperial Aramaic). For example *gimel* represents the letters *dāleth*, and *yōdḥ*, but also the corrupt forms of *bēth*, *zayin*, and *kaḥp*. Combinations of these letters, accidental reduplications, and the fact that two *yōdḥ* resemble both *'āleph* and *sāmekh*, create further difficulties.

The chief characteristic of Pahlavi is that, in spite of its being phonetically purely Iranian, it mixes Semitic and Iranian words in its orthography. From an early period, writers of Pahlavi used Aramaic words as ideograms, that is, they no longer pronounced or even thought of them as Semitic words but as familiar shapes or signs only, written to convey Iranian equivalents. The Semitic words are not ideograms in the sense of Sumeric or Chinese symbolic characters, but are written with a consonantal alphabet, and so may be better termed 'heterograms,' and Iranian phonetic spellings 'eteograms' (Klíma 1968: 28). By

convention, modern scholars transliterate ideograms in upper case type, phonetic spellings in lower case. Book Pahlavi is written in an almost equal mixture of Iranian eteograms (e.g., *pyt'k'* = *paydāg* 'revealed,' *gwpt'* = *guft* 'said,' – but cf. also *YMRRWNt* = *guft*) and Aramaic heterograms (e.g., *GBRA* 'man' was read and pronounced as Iranian *mard*, and *YWM* 'day' renders Iranian *rōz*). Usually verbal stems are written heterogramatically, with Iranian phonetic inflexional endings, e.g., *YHWWNy't* = *bawēd* 'it is.' Occasionally a word can be read as both an ideogram or as a phonetic spelling, e.g., *TWB/did* is like *tang*, and *LHYK/dūr* is identical to *Ihyk/rahīg*. Problems of haplography and dittography blemish the manuscripts and make reading difficult: since Pahlavi was no longer a vernacular language in Islamicized Iran, nor was its script used outside the Zoroastrian religious texts after about 700 AD, copyists often did not understand what they were writing.

The greater part of Sasanian literature in Pahlavi was in fact secular poetry, but this has not survived in its original form, having been translated to suit Islamic tastes into New Persian, notably in the *Shāhnāma* of Ferdāusi (c. 935–c. 1020), in Arabic script. Those who remained faithful to Zoroastrianism after the Islamic conquest (seventh century AD) continued to use Pahlavi to preserve their scriptures and religious lore in the archaic orthography which kept them obscure to all except Zoroastrians. They are of interest to the historian of religions because of the richness of their theological and mythological content, but, with a few exceptions, they are of limited literary merit. Pahlavi was a sonorous and robust language, which is recognizably the source of the characteristic mellifluous qualities of New Persian, even after centuries of arabicization.

Bibliography

- Boyce M (1968). Middle Persian literature. In Spuler B (ed.) *Iranistik*. Leiden: E J Brill.
- de Menasce J P (1983). Zoroastrian Pahlavi writings. In Yarshater E (ed.) *Cambridge History of Iran*, vol. 3. Cambridge: Cambridge University Press.

Klíma O (1968). The Middle Persian era. In Rypka J (ed.) *History of Iranian Literature*. Dordrecht: D. Reidel.
 MacKenzie D N (1967). Notes on the transcription of Pahlavi. *Bulletin of the School of Oriental and African Studies* xxx, 17–29.

MacKenzie D N (1983). *A Concise Pahlavi Dictionary*. Oxford: Oxford University Press.
 West E W (1904). Pahlavi literature. In Geiger W & Kuhn E (eds.) *Grundriß der Iranischen Philologie*, vol. II. Strassbourg.

Palenquero

A Schwegler, University of California, Irvine, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Palenquero is a Spanish-lexicon creole (*see Pidgins and Creoles*) spoken in the village of El Palenque de San Basilio, Colombia. Located 60 km inland from the former slave trade center of Cartagena de Indias (*see Figures 1 and 2*), this ethnically homogenous Afro-Hispanic community is inhabited by descendants of runaway African slaves who, around 1700, established their first *palenques* (primitive fortifications) in the interior of the Caribbean coast. Palenquero is unique in that it is the only known Spanish-based creole on the South American mainland (Lipski and Schwegler, 1993).

Until the early 1990s, the Palenqueros lived in relative cultural and geographic isolation (Schwegler, 1996, 1998; Schwegler and Morton, 2003), which significantly contributed to the preservation of the local creole, although historically, they have always maintained some contact with the outside world; Palenquera women in particular visit nearby towns and Cartagena on a regular basis, where they generally sell and trade locally produced goods. However, this situation changed rather dramatically in the 1990s and beyond, when word of the existence of this ‘African’ village in the hinterland of Cartagena spread rapidly in academic circles. In Colombia and elsewhere, recent documentaries about Palenquero culture have contributed to the relative fame of the community. The stream of visitors to El Palenque has, however, subsided of late because guerrilla activities in nearby areas have made local travel rather risky for outsiders.

The Palenquero community has been bilingual (Spanish/creole) for at least two centuries. Starting around 1970, however, adolescents in particular began to shun the use of ‘Lengua,’ the local name of the creole. Today Spanish monolingualism is the norm among the younger generations, though many still possess a passive knowledge of the local vernacular. In recent years (1990s onward), there has been a growing awareness of ‘*negritud*’ (black pride) among both Palenqueros and Afro-Colombians. This has led to

modest institutional and political support to counter the loss of the Palenquero creole. The local elementary school, for instance, now offers some courses in Lengua, and a few adolescents have been attempting to devise ‘official’ spelling conventions for the creole. Also, some Palenqueros have begun to consciously adapt sub-Saharan vocabulary (including words such as *Bantú*) so as to identify, strengthen, and celebrate what is, in their view, African in their heritage.

Almost the entire Palenquero lexicon is derived from Spanish, and the phonetic distance between most creole and (Caribbean) Spanish words is relatively minor. For example, representative Spanish/creole vocabulary sets are: *manol/mano* ‘hand,’ *hom-brel/ombe* ‘man,’ *dedollelo* ‘finger,’ *senti(r)/sindí* ‘to feel,’ *agarra(r)/angalá* ‘to grab, to hold.’ But despite such close lexical correspondences, Palenquero and Spanish are scarcely mutually intelligible. Differences in grammar are the main reason for this unintelligibility.

Historically, a key feature of local language use has been intense and very rapid code switching (not to be



Figure 1 Location of El Palenque.

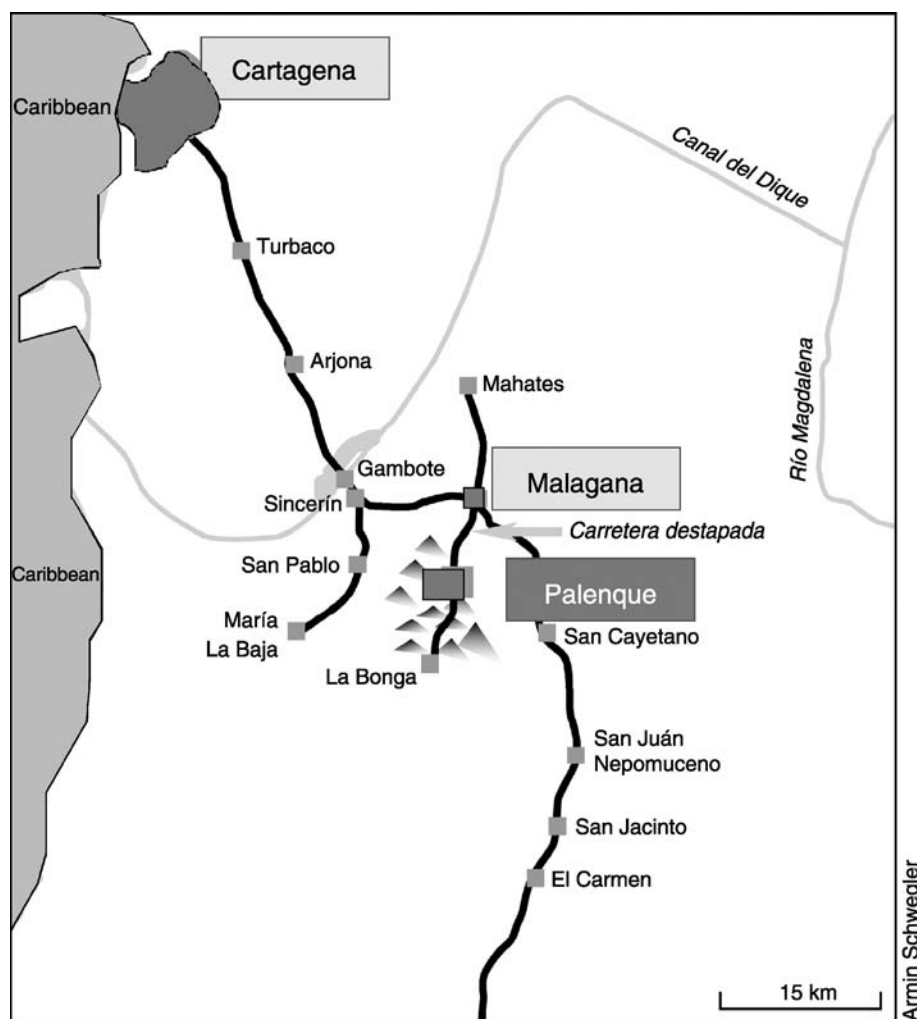


Figure 2 Cartagena/Palenque area.

confused with code mixing). The following example illustrates how speakers tend to switch language – often multiple times – within a single utterance (segments within angle brackets are in regional Spanish):

Muhé mi <no quiere> komblá- mi pekao,
 woman my not want.3s buy- me fish
 <a meno que yo vaya> ku ele.
 unless I go.PRES.SUBJ. with him/her
 ‘My wife doesn’t want to buy me fish unless I go (buy it) with her.’

It is now clear that the Kikongo (Kongo) language, spoken in central west Africa (see **Figure 3**), played a pivotal role in the genesis of Palenquero. As in Cuba (Fuentes and Schwegler, 2005), in El Palenque Bakongo slaves seem to have passed down their African language for several generations, either as a ritual code or as a full-fledged everyday means of communication. Scholars have also been able to determine that Bantu (rather than west African) fugitives must have had the most profound impact on

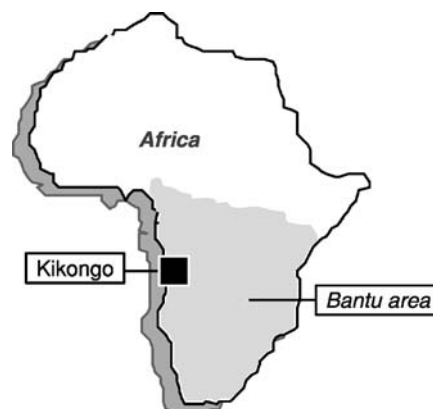


Figure 3 Location of the Kikongo language.

El Palenque’s early language and culture (Schwegler, 1996, 2002, forthcoming).

Detailed information about the linguistic and cultural history of El Palenque can be found in Moñino and Schwegler (2002), Schwegler (1998,

2002, 2006), Schwegler and Morton (2003), and Schwegler and Green (2007). These studies also list earlier publications on the topic, including Friedemann and Patiño Rosselli (1983), still the most solid description of Palenquero to date. Importantly, the volume contains the only substantial corpus of Palenquero texts (readers should be aware, however, that the authors omitted to differentiate code switches from Lengua to Spanish).

Bibliography

- Friedemann N S de & Patiño Rosselli C (1983). *Lengua y sociedad en El Palenque de San Basilio*. Bogota: Instituto Caro y Cuervo.
- Fuentes J & Schwegler A (2005). *Lengua y ritos del Palo Monte Mayombe: dioses cubanos y sus fuentes africanas*. Frankfurt: Vervuert/Madrid: Iberoamericana.
- Lipski J & Schwegler A (1993). 'Creole Spanish and Afro-Hispanic.' In Green J N & Posner R (eds.) *Bilingualism and linguistic conflict in Romance*. Berlin: Mouton de Gruyter. 407–432.
- Moñino Y & Schwegler A (eds.) (2002). *Palenque, Cartagena y Afro-Caribe: historia y lengua*. Tübingen: Max Niemeyer.
- Schwegler A (1996). 'Chi ma nkongo': *lengua y rito ancestrales en El Palenque de San Basilio (Colombia)* (2 vols.). Frankfurt: Vervuert/Madrid: Iberoamericana.
- Schwegler A (1998). 'Palenquero.' In Perl M & Schwegler A (eds.) *América negra: panorámica actual de los estudios lingüísticos sobre variedades criollas y afrohispanas*. Frankfurt: Vervuert/Madrid: Iberoamericana. 220–291.
- Schwegler A (2002). 'On the (African) origins of Palenquero subject pronouns.' *Diachronica* 19(2), 273–332.
- Schwegler A (2006). 'Bantu elements in Palenque (Colombia): anthropological, archeological and linguistic evidence.' In Havisser J B & MacDonald K (eds.) *African re-gensis: confronting social issues in the diaspora*. London: UCL Press.
- Schwegler, A & Green K (2007). 'Palenquero (Creole Spanish).' In Holm J & Patrick P (eds.) *Comparative creole syntax*. London: Battlebridge.
- Schwegler A & Morton T (2003). 'Vernacular Spanish in a microcosm: *Kateyano* in El Palenque de San Basilio (Colombia).' *Revista Internacional de Lingüística Iberoamericana* 1, 97–159.

Pāli

J W Gair, Cornell University, New York, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Pāli (also Pāli and Pali) is an early Middle Indo-Aryan (MIA) language, or Prakrit. It is the text and ritual language of Theravada, or southern, Buddhism, the dominant school in Sri Lanka, Burma, Cambodia, and Thailand. It is of particular importance as the language in which the basic teachings of Buddhism have been preserved, especially in the collection known as the *Tipiṭaka* (literally, 'three baskets'), which are held to contain the Buddha's own pronouncements. Virtually all of the extensive Pāli literature is thus Buddhist in nature or origin, and the language is not spoken except in recitation and as an occasional vehicle of communication for monks of different languages.

The date and place of origin of Pāli have been subjected to considerable scholarly debate through the years, and the position that one accepts may not unnaturally be colored by belief as to the authenticity of the canonical texts as the word of the Buddha as spoken by him. By tradition, especially in Sri Lanka, the language, as the vehicle of the Buddha's preaching, would date from his time (7–6 century B.C.E.) and be identified with Māgadhī, the language of

Magadha, the northeastern India kingdom in which he primarily preached. His date, however, varies somewhat in different traditions, and scholars in both India and the West have argued for progressively later dates – some as late as the 4th century B.C.E. Also, numerous scholars have pointed out that Pāli not only does not share many of the distinctive characteristics of the Māgadhī Prakrit as shown in later inscriptions, primarily those of the 3rd century B.C.E. Emperor Asoka (Sanskrit *Aśoka*) (264–227 B.C.E.), but it does in fact share important features, such as noun inflections, of the western inscriptions. Thus, Pāli does not appear to represent any single MIA dialect but to be a literary language that incorporated features from several dialects in the course of its development.

The canonical texts were transmitted orally for a number of centuries and were collected and codified in three main councils: first at Rājagaha (Sanskrit *Rājagṛha*) shortly after the death of the Buddha, and then at Vesālī (Sanskrit *Vaiśālī*), about a century later. The third, at Pāṭaliputta (Sanskrit *Pāṭaliputra*), under Emperor Aśoka. There, the canon as we know it was essentially completed and formalized, the Theravāda school founded, and the decision taken to send missions abroad made, including the mission that brought the doctrine to Sri Lanka through the monk Mahinda. The generally accepted view is that the

Table 1 The Pāli sound inventory**Vowels**

a ā, i, ī, u, ū, e, o

Consonants

Velars: k, kh, g, gh, ŋ

Palatals: c, ch, j, jh, ñ

Cerebrals (Retroflex): ṭ, ṭh, ḍ, ḍh, ṇ

Dentals: t, th, d, dh, n

Labials: p, ph, b, bh, m

Resonants: y, r, l, ṽ

Spirants: s, h

canon was reduced to writing only in the 1st century B.C.E. at the Aluvihāra in Sri Lanka.

Pāli has no special alphabet of its own but is written in several scripts, depending on the country and the intended audience. Thus, it commonly appears in Sinhala script in Sri Lanka, in Devanāgarī in India, and in Burmese, Cambodian, and Thai in those countries. In the West, and where it is intended for an international audience, it is commonly written in the Roman alphabet with some diacritics.

The Pāli system of sound elements is given in Table 1. It is, of course, represented differently in different scripts. The usual alphabetical order can be read by taking each row in turn, from top to bottom, and some manuscript traditions include a ‘pure nasal’ symbol, transliterated as <ṁ>, occurring between the vowels and consonants. It represents ṇ at the end of words, but before a consonant assimilates to it.

This is essentially the same inventory of elements as Sanskrit, though there were intervening changes that gave Pāli, like Prakrits in general, a reduced inventory. Among the most important were the following: Sanskrit vocalic *r* was lost, becoming *i*, *a*, or *u*. The three Sanskrit sibilants were merged as *s*, and all final nasals as *ṇ*. Long vowels were shortened in checked syllables, and this extended to Sanskrit *e* and *o* (always long in Sanskrit, but in Pāli allophonically short before consonant sequences). Thoroughgoing changes applied to consonant sequences (clusters). These were numerous and complex, and there were variations and exceptions owing to dialect admixture and the long oral and textual history. But generally, some initial clusters were simplified, sometimes with the addition of a prothetic vowel, and internal clusters were assimilated internally, yielding many geminates, with sibilants becoming aspiration in some combinations. Thus Sanskrit *strī* is Pāli *itthī* ‘woman’, and Sanskrit *asti* is Pāli *atthi* ‘is’. Sanskrit *svarga* is Pāli *sagga*, ‘heaven’, Sanskrit *dharma* is Pāli *dhamma* ‘doctrine’ (and many other meanings), and Sanskrit *prajñā* is *paññā*. Sanskrit *akṣi* is *akkhi* ‘eye’ (also *acchi*, probably showing dialect admixture). Sanskrit *lakṣaṇa* is Pāli *lakkhaṇa* ‘feature’. Sanskrit *mārga* is

Pāli *magga* ‘way, path’, showing long vowel reduction, and the common assimilation, but *dīgha* ‘long’, Sanskrit *dīrgha*, shows an alternate development: simplification of the cluster and retention of vowel length.

In morphology, Pāli remained an inflectional language, but there were numerous changes from Sanskrit in grammatical categories and forms, including simplifications and confluences. Thus, in nouns many case affixes have fallen together; in the verb, the Sanskrit past vs. aorist distinction has virtually disappeared, with a new past based on the aorist, and in both nouns and verbs the dual is gone.

Pāli basic word order is verb-final, i.e., Subject-Object-Verb, as in (1):

- (1) bhikkhu cittam pagganhāti
 monk mind-accusative uplifts
 ‘The bhikkhu uplifts the mind.’

However, there is much variation for pragmatic effects such as foregrounding, and in some types of existential and interrogative sentences, as in (2) and (3):

- (2) atthi koci satto, yo imamhā
 be any being that this
 PRES-3sg (REL) ABL
 kāyā aññaṃ kayam saṃkamati?
 body other body transmigrate
 ABL ACC ACC PRES-3sg
 ‘Is there any being that migrates from this body to another body?’
- (3) natthi satto yo evaṃ saṃkamati.
 not-be being that thus transmigrate.
 PRES-3sg (REL) PRES-3sg
 ‘There is no being that so transmigrates.’

Pāli also uses the correlative relative construction common in Indo-Aryan languages, as in (4), though there are also ‘simple’ relatives, as (2) and (3) have exhibited:

- (4) yaṃ jānāmi taṃ bhāṇāmi
 what know that speak
 CORREL PRES-1sg PRES-1sg
 ‘I say what I know.’

Pāli literature can be divided into two sets: canonical and non-canonical. Canonical texts are generally those regarded as the actual teachings of the Buddha, though there is some difference in what is included in the canon in different countries. The most widely known traditional classification of the canon is the *Tiṭṭaka* (‘Three Baskets’), by which there are three main divisions or Pitakas, the *Sutta*, *Vinaya*, and *Abhidhamma*. These can be generally characterized as follows.

- I. The *Sutta Piṭaka* contains the *Dhamma* proper (General teachings of the Buddha), and it is sometimes referred to as such. It contains five *Nikāyas*, or collections of *suttantas* (Dialogues of the Buddha), defined and arranged essentially by their form, as follows:
 - a. The *Dīgha Nikāya* ('Long' Collection) contains the longest suttas (Sanskrit *sūtra*).
 - b. The *Majjhima Nikāya* ('Middle' Collection) contains suttas of middle length.
 - c. In the *Samyutta Nikāya* ('Linked' or 'Grouped' Collection), the suttas are arranged by topic. It is this collection that contains the Buddha's first sermon, the *Dhammacakkapavattanasutta*.
 - d. The *Aṅguttara Nikāya* (or The 'Gradual', or 'By one limb more' Collection), in which the sections are arranged in ascending order according to numbers that figure in the texts themselves.
 - e. The exact contents of the *Khuddaka Nikāya* ('Short' or 'Small' Collection) vary somewhat between Sri Lanka, Burma, and Thailand. It includes the widely known *Dhammapada*. It also contains the *Jātaka* verses, but only the verses, not the birth stories connected with them, are canonical; the stories are considered to be commentarial. It also includes the hymns of the monks and nuns (*Theragāthā* and *Therīgāthā*) along with a number of other works such as the *Suttanipāta* and some works that might be loosely categorized as 'prayer books'.
- II. The *Vinaya Piṭaka* dealing with Monastic Discipline.
- III. The *Abhidhamma Piṭaka*. Scholastic and partially metaphysical in nature, it contains much philosophical treatment of the Buddha's teachings. It is generally considered the most difficult of the texts, so that a mastery of it is highly valued by Buddhist scholars.

There is another traditional classification of the canon into five divisions (*Nikāyas*). These are the five divisions of the *Sutta Piṭaka* of the *Tipiṭaka*, with the *Abhidhamma* and the *Vinaya* folded into the *Khuddaka Nikāya*.

In addition to the above, there is the *Mahāparitta*, a text recited by monks at *paritta* (Sinhala *pirit*) ceremonies invoking the auspiciousness and protection of the *Dhamma*.

In addition to the canonical texts, there is a considerable body of literature in Pāli, continuing up to the present time, and much of it is commentarial literature or chronicles. The remainder includes various types of works, including narrative and

instructional works and some grammars. In addition, there are a number of inscriptions, most of them in Southeast Asia.

The commentarial literature in Pāli continued over many centuries, but the most famous commentaries, or *aṭṭhakathās*, were by a monk named Buddhaghosa, in the 5th century A.D. He was born in South India but wrote his commentaries in Sri Lanka, apparently basing much of his work on earlier Sinhala commentaries subsequently lost. He also authored the famous *Visuddhimagga* 'Path of Purification', a compendium of Buddhist doctrine. As mentioned earlier, the well-known *Jātaka* stories are actually commentaries on the *Jātaka* verses that are included in the canon, and this *Jātakaṭṭhakatā* has also been attributed to Buddhaghosa. In addition to the commentaries, there are other forms of commentarial literature, including *tīkā*s, subcommentaries on the commentaries.

The Chronicles include the *Dīpavaṃsa* (4th or early 5th century A.D.) and the *Mahāvaṃsa* (probably the early 6th century), and they present the history of Sri Lanka from a Buddhist-Monastic perspective. These chronicles were continued by the *Cūlavāṃsa*, which continued until the arrival of the British in Sri Lanka. In fact, they are being continued even today.

Among the remaining works, the *Milindapañhā* (sometimes in the singular *Milindapañho*) 'Questions of King Milinda' is particularly appealing. It dates from before Buddhaghosa's commentaries, may have been translated from Sanskrit, and was itself translated into Chinese. It consists of a series of dialogues between two people: King Milinda (Greek Menander), a second century king of a Graeco-Bactrian kingdom remaining from Alexander the Great's incursions into what is now Afghanistan and the northwest Indian subcontinent, and Nāgasena, a learned monk, who expounds Buddhist doctrine in answer to the King's questions. The penetrating nature of the King's questions and the clarity and wit of Nāgasena's answers and explanations make this still a lively as well as instructive introduction to Buddhist doctrine, and one that is accessible to the student at a fairly early stage in learning Pāli.

There is now a sizeable and growing amount of material in and on Pāli on the World Wide Web. The Pali Text Society, founded in 1881, has published many texts and translations in roman script. Its website has information on the available ones. Fifty-eight volumes of the *Tipitaka* were published, in Sinhala script with a Sinhala translation, as the Buddha Jayanti Tripitaka Series under the patronage of the government of Sri Lanka (Ceylon) during the 1960s and 1970s. The Pāli text in roman transcription, along with some paracanonical and other texts, has been made available online as a free public-domain

edition by the Sri Lanka Tipitaka Project in association with the *Journal of Buddhist Ethics*.

Bibliography

- Buddhadatta A P Mahāthera (1955). *English-Pali dictionary*. Oxford: The Pali Text Society. (1989 reprint: Delhi: Motilal Banarsidas).
- Buddhadatta A P Mahāthera (1989). *Concise Pali-English dictionary*. Delhi: Motilal Banarsidas (reprint of 1957 Colombo Apothecaries printing).
- Geiger W (1994). *A Pāli grammar*. (1916). Ghosh B and Norman K R (trans. and ed.) (1943). Oxford: The Pali Text Society.
- Gair J W & Karunatilake W S (1998). *A new course in reading Pāli: entering the word of the Buddha*. Delhi: Motilal Banarsidas.
- Norman K R (1983). *Pali Literature*. Wiesbaden: O. Harrassowitz.
- Oberlies T (2001). *Pāli-a grammar of the language of the Theravada Tipiṭaka*. Berlin and New York: De Gruyter.
- Oberlies T (2003). 'Asokan Prakrit and Pali.' In Cardona G & Jain Dh (eds.) *The Indo-Aryan languages*. London and New York: Routledge. 161–203.

Perniola, Vito S J (1997). *Pali Grammar*. Oxford: The Pali Text Society.

Rhys Davids T W & Stede W (1921–1925). *Pali-English Dictionary*. Oxford: The Pali Text Society. (1975 reprint, New Delhi: Munshiram Manoharlal.)

Warder A (1963). *Introduction to Pali*. Oxford: The Pali Text Society.

Relevant Websites

- <http://www.palitext.com> – Information on the available texts and translations in roman script is available here.
- <http://dsal.uchicago.edu/dictionaries/pali/index.html> – The Pali Text Society dictionary can be accessed here.
- <http://jbe.gold.ac.uk/palicanon.html> – The Pāli text in roman transcription of 58 volumes of the Tipitaka is available in a free public-domain edition here.
- <http://www.tipitaka.org> – The Vipassana Research Institute Tipitaka Project offers various texts.
- <http://www.accesstoinsight.org> – This site provides useful links and information maintained by John Bullitt.
- <http://www.metta.lk> – Tipitaka texts with Sinhala and English translation (at present Windows only), and other materials including dictionary, grammar, and lessons.

Panoan Languages

D W Fleck, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

The Panoan language family is composed of approximately 30 known languages spoken in the western Amazon basin, in eastern Peru, western Brazil, and northern Bolivia. Of these, only about 20 are still spoken today and most are in danger of extinction. Additionally, there are several uncontacted groups in westernmost Brazil suspected to be Panoans (Erikson, 1994). There are currently 40 000–50 000 speakers of Panoan languages.

History and Culture

Archeological evidence suggests that the ancestral homeland of the Panoans was in northern Bolivia and that they migrated northward around 300 A.D. (Myers, 1990: 99). In past centuries, factions of many Panoan groups were reduced at Jesuit and Franciscan missions in Peru. Currently, Panoans occupy a fairly continuous territory and are relatively homogeneous

linguistically and culturally (Erikson, 1992). Traditional subsistence, still practiced today by most groups, consists mainly of slash-and-burn horticulture, hunting with bow and/or blowgun, and fishing.

Classification

The Panoan languages were recognized early on by Jesuit missionaries to be closely related (e.g., in a 1661 letter by Father Francisco de Figueroa; Figueroa *et al.*, 1986: 214). The first formal demonstration that Panoan languages constitute a linguistic family was in 1888 by Raoul de la Grasserie, based on a comparison of eight word lists of Panoan languages/dialects collected by European explorers earlier that century (Grasserie, 1890). The family was named after the now-extinct Pano language (also known as Panobo 'giant armadillo people' or Wariapano). There is still no authoritative subclassification of the Panoan family available; see Valenzuela (2003b) for an evaluation of past subgroupings of the family. It has been claimed that the Panoan family is undoubtedly related to the Tacanan family (e.g., Suárez, 1973), though today not all Panoan scholars accept this as certain.

Phonology

Loos (1999) reconstructs the following phoneme inventory for proto-Panoan: p, t, k, ʔ, ts, tʃ, s, ʃ, ʂ, β, ɾ, m, n, w, j, h, a, i, ɪ and o. Most languages have rhythmic stress, where every other syllable in a word is stressed.

Morphology

Panoan languages are primarily suffixing and could be called highly synthetic due to the potentially very long words (up to about 10 morphemes), but the typical number of morphemes per word in natural speech is not large. It is the large number of morphological possibilities that is striking about Panoan languages. For example, up to about 130 different verbal suffixes express such diverse notions as causation, direction of movement, evidentiality, emphasis, uncertainty, aspect, tense, plurality, repetition, etc. Panoan languages are all morphologically ergative, with an ergative case marker that also marks instrumental and genitive cases, and in some languages also locative and/or vocative. Complex and sometimes obligatory systems of evidentiality (Valenzuela, 2003a) and body part prefixation (Fleck, 2006) are two further notable features of Panoan morphology.

Syntax

Panoan languages have the rare and interesting property of ‘transitivity agreement,’ where various parts of the grammar (including adverbs, suffixes, and enclitics) vary depending on whether the matrix verb is transitive or intransitive. Panoan discourse is characterized by ‘clause chaining’ (or ‘switch reference’): up to about 10 clauses can be linked together using suffixes that mark argument coreference (e.g., same subject, object = subject) and temporal/logical relations (e.g., ‘while,’ ‘after,’ ‘in order to’) between subordinate and matrix clauses. Panoan languages are some of the few languages in the world where both nonsubject arguments of bitransitive verbs such as *give* are grammatically identical. See Sparing-Chávez (1998), Valenzuela (1999, 2003b), Faust and Loos (2002), and Fleck (2003) for modern descriptions of these and other Panoan grammatical phenomena.

Lexicon and Ethnolinguistics

Some Panoan groups have a taboo that prohibits mention of a deceased person’s name and nicknames, otherwise the dead person’s spirit may cause harm to the family of the person that pronounces his/her name

out loud. The name taboo also prohibits mentioning words judged to sound like the deceased’s name or nicknames. Languages such as Matses seem to have an unusually high rate of lexical replacement, probably due at least in part to name taboo. Other ethnolinguistic features of interest in Panoan languages are parent-in-law avoidance speech in Shipibo-Konibo (Valenzuela, 2003b) and elaborate rain forest habitat classification nomenclature (e.g., Matsés has 47 terms for types of rain forest; Fleck and Harder, 2000).

Bibliography

- Erikson P (1992). ‘Uma singular pluralidade: a etno-história Pano.’ In Carneiro da Cunha M (ed.) *História dos índios no Brasil*. São Paulo: Companhia das Letras. 239–252.
- Erikson P (1994). ‘Los Mayoruna.’ In Santos F & Barclay F (eds.) *Guía etnográfica de la alta Amazonía 2*. Quito: Flasco-Sede. 1–127.
- Faust N & Loos E E (2002). *Gramática del idioma Yaminahua*. Lima: Instituto Lingüístico de Verano.
- Figueroa F de, Acuña C de et al. (1986). *Informes de Jesuitas en el Amazonas 1660–1684*. Iquitos, Peru: CETA/IIAP.
- Fleck D W (2003). ‘A grammar of Matses.’ Ph.D. diss., Rice University, Houston.
- Fleck D W (2006). ‘Body-part prefixes in Matses: derivation or noun incorporation?’ *International Journal of American Linguistics* 72(1), 59–96.
- Fleck D W & Harder J D (2000). ‘Matses Indian rain-forest habitat classification and mammalian diversity in Amazonian Peru.’ *Journal of Ethnobiology* 20, 1–36.
- Grasserie R de la (1890). ‘De la famille linguistique Pano.’ *Congrès International des Américanistes: compte-rendu de la septième session, Berlin 1888*, 438–449.
- Loos E E (1999). ‘Pano.’ In Dixon R M W & Aikhenvald A Y (eds.) *The Amazonian languages*. Cambridge: Cambridge University Press. 227–249.
- Myers T P (1990). *Sarayacu*. Lincoln: University of Nebraska.
- Sparing-Chávez M W (1998). ‘Interclausal reference in Amahuaca.’ In Derbyshire D C & Pullum G K (eds.) *Handbook of Amazonian languages 4*. Berlin: Mouton de Gruyter. 443–485.
- Suárez J A (1973). ‘Macro-Pano-Tacanan.’ *International Journal of American Linguistics* 39, 137–154.
- Valenzuela P (1999). ‘Adverbials, transitivity, and switch-reference in Shipibo-Konibo.’ *Chicago Linguistic Society 35: The Panels*, 355–371.
- Valenzuela P M (2003a). ‘Evidentiality in Shipibo-Konibo, with a comparative overview of the category in Panoan.’ In Aikhenvald A Y & Dixon R M W (eds.) *Studies in evidentiality*. Amsterdam: John Benjamins. 33–61.
- Valenzuela P M (2003b). ‘Transitivity in Shipibo-Konibo grammar.’ Ph.D. diss., University of Oregon, Eugene.

Papiamentu

T Sanchez, University of Pennsylvania, Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Papiamentu is a Creole language spoken on Aruba, Bonaire, and Curaçao in the Caribbean. Over 175 000 islanders (about 75% of residents) speak the language natively, and many immigrants learn it as a second language. It is widely used in both public and private domains, for artistic and practical purposes, and is included in secondary education. The earliest surviving written example is a personal letter from 1775, and many 19th-century texts also exist, including translated religious documents and news articles originally written in the Creole. Today, most Papiamentu speakers have varying levels of competence in Dutch (the official language), Spanish, and English.

Origins Researchers do not agree on whether Papiamentu was formed around Spanish (Maduro, 1966) or Portuguese (Maurer, 1986; Goodman, 1987; Martinus, 1996). Proponents of the Spanish origin suggest that the creole formed during the 16th century from contact between the Spanish and Caquetio Indians. But it is more likely that Papiamentu was formed during the latter half of the 17th century from the speech of Portuguese-speaking Jewish merchants and African slaves, with influence from Dutch colonists, Spanish traders, and native Caquetios. Today most lexical items resemble Portuguese or Spanish, and to recognize both influences, we say that Papiamentu has an ‘Iberian’ lexical base.

Orthography Papiamentu has two orthographic traditions: Aruba prefers an etymological system, while Curaçao and Bonaire follow a phonological system. The phonological system is used here.

Phonology The vowel inventory of Papiamentu is a, ε, e, ø, ɔ, o, i, y, u. The front round vowels were introduced via Dutch lexical items; the mid round vowels are found in the Portuguese and Dutch lexicons. Consonants are p, b, t, d, k, g, s, z, ʃ, ʒ, h, tʃ, dʒ, m, n, ŋ, l, r, w, j. Lexical tone, stress, and sandhi phenomena are part of Papiamentu’s prosodic structure.

Morphology Papiamentu has a few productive affixes, including *-mentu* ‘the act of,’ from Spanish *-miento* (i.e., *distribimentu* ‘(the act of) wasting,’ *kapmentu* ‘cutting,’ and *kèchmentu* ‘catching’); *-dó* ‘person who’, from Spanish *-dor* (i.e., *wardadó* ‘keeper, guard’, lit. ‘person who guards’; *trahadó* ‘worker’;

huurdó ‘tenant’) (Dijkhoff, 1993); plural marker *-nan*; and gerundive and progressive marker *-ndo* from Spanish *-ndo* (Sanchez, 2002, 2005). Borrowed morphemes which are not yet completely integrated may be sensitive to etymology. For example, *-ndo* is productive with Iberian verbs, and though it is attested with Dutch verbs, such usage is unacceptable for most speakers. Past participles are formed by shifting stress to the final syllable, but some Dutch-origin verbs take *he-* (as in Dutch) instead. Past participles may be semantically extended as nouns (e.g., *kasa* ‘marry’ → *kasá* ‘married’ → *kasá* ‘spouse’).

Syntax The basic word order of Papiamentu is SVO. It is neither pro-drop like Spanish and Portuguese, nor V2 like Dutch; pronominal objects cannot be moved to preverbal position, and there is no wh-movement. As in many creoles, tense, mood, and aspect are indicated by preverbal markers:

| | |
|---------------|-------------------|
| <i>ta</i> | imperfective |
| <i>tabata</i> | past imperfective |
| <i>a</i> | perfective |
| <i>lo</i> | future |
| <i>sa</i> | habitual |

(based on the analysis in Andersen, 1990).

Papiamentu also has a passive voice, composed of a preverbal marker, a passivizing verb (*ser*, *wordu*, or *keda*), and a past participle (e.g., *ta wordu skuchá* ‘is heard’).

Bibliography

- Andersen R W (1990). ‘Papiamentu tense-aspect, with special attention to discourse.’ In Singler J (ed.) *Pidgin and creole tense-mood-aspect systems*. Amsterdam: John Benjamins. 59–96.
- Dijkhoff M (1993). *Papiamentu word formation: a case study of complex nouns and their relation to phrases and clauses*. Amsterdam: University of Amsterdam Centrale Drukkerij.
- Fouse G C (2002). *The Story of Papiamentu: A Study in Slavery and Language*. Lanham, MD: University Press of America.
- Goodman M (1987). ‘The Portuguese element in the American creoles.’ In Gilbert G (ed.) *Pidgin and creole languages: essays in memory of John E. Reinecke*. Honolulu: University of Hawai’i Press. 361–405.
- Howe K (1994). *Papiamentu reader*. Kensington, MD: Dunwoody Press.
- Kouwenberg S & Murray E (1994). *Papiamentu*. Munich: Lincom Europa.
- Maduro A J (1966). *Procedencia di palabranan Papiamentu i otro anotacionnan*. Curaçao, Netherlands Antilles: privately printed.

- Maduro A J (1991). *Papiamentu: indagando i ilustrando*. Curaçao, Netherlands Antilles: Drukkerij Scherpenheuvel N.V.
- Martinus E F (1996). 'The kiss of a slave: Papiamentu's West-African connections.' Ph.D. thesis, University of Amsterdam.
- Maurer P (1986). 'El origen del Papiamentu (desde el punto de vista de sus tiempos gramaticales).' *Neue Romania* 4, 129–149.
- Rivera-Castillo Y (1998). 'Tone and stress in Papiamentu: the contribution of a constraint-based analysis to the problem of creole genesis.' *Journal of Pidgin and Creole Languages* 13(2), 297–334.
- Romer R (1977). 'Polarization phenomena in Papiamentu.' *Publikaties van het Instituut voor Algemene Taalwetenschap* 17, 69–79.
- Salomon H P (1982). 'The earliest known document in Papiamentu contextually reconsidered.' *Neophilologus* 66, 367–376.
- Sanchez T (2002). 'The interacting influences of Spanish and English on the creole Papiamentu.' *University of Pennsylvania Working Papers in Linguistics* 8.3, 235–247.
- Sanchez T (2005). 'Constraints on structural borrowing in a multilingual contact situation.' Ph.D. thesis, University of Pennsylvania.
- Wood R (1971). 'The hispanization of a creole language: Papiamentu.' *Hispania* 55, 857–864.
- Wood R (1972). 'Dutch Syntactic Loans in Papiamentu.' *Revue des Langues Vivantes* 38, 635–647.

Papuan Languages

A Pawley, Australian National University, Canberra, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

'Papuan' is a collective name for a number of language families and genetic isolates that have in common two characteristics: (a) they are indigenous to a region sometimes called the New Guinea area, comprising New Guinea and neighboring island groups extending from Timor, Alor and Pantar, and Halmahera in the west to the Solomon Islands in the east; and (b) they do not belong to the vast Austronesian family, which dominates Island Southeast Asia and the archipelagoes of the southwest and central Pacific but is only patchily represented in New Guinea itself. (The term 'family' will be used here exclusively to refer to linguistic groups of the highest genealogical order, not to subgroups.)

The hub of the Papuan-speaking region is the large island of New Guinea, which is about the size of Germany but contains about 900 mutually unintelligible languages, over 700 of which are Papuan. According to the most recent classifications, some 18 Papuan families and several isolates are represented on the New Guinea mainland (see **Figure 1**). Two of the New Guinea-based families also have members in Alor, Pantar, and Maluku in Indonesia and in East Timor. Another five, possibly six, families and several isolates are found in the arc of islands extending from New Britain to the Solomons (see **Figure 2**). Whereas Austronesian languages arrived in Melanesia from the west within the past 3500

years (Spriggs, 1997), the Papuan families almost certainly represent continuations of linguistic stocks that have been in this region for much longer than this. There is no convincing evidence that any of the Papuan families have relatives outside of the New Guinea area.

About three million people speak Papuan languages. Most have fewer than 3000 speakers. The seven largest language communities are Enga (about 200 000) and Medlpa [Melpa] (150 000), of the highlands of Papua New Guinea, and Western Dani (150 000) and Lower Grand Valley Dani (130 000) of the highlands of Irian Jaya (Papua). The small size of language communities reflects the extreme political fragmentation that is characteristic of the New Guinea area; peoples were traditionally subsistence farmers or foragers and until colonial times political groups seldom exceeded a few hundred people. In postcolonial times the main regional lingua francas in the Papuan-speaking regions have been English and Tok Pisin in Papua New Guinea, English and Pijin in the Solomon Islands, and Malay in Indonesia. No Papuan language has the status of a national or even a provincial language. While most Papuan languages are still vibrant in their local communities, their small size and lack of wider status mean that their long-term prospects of survival are poor.

Foley (1986) gives an excellent overview of Papuan languages and linguistics up to the mid-1980s; Foley (2000) reviews more recent work. Carrington (1996) is a near exhaustive bibliography of linguistic research up to 1995 and Laycock and Voorhoeve (1971) is a thorough history of early research. *Language atlas of the pacific area* (Wurm and Hattori, 1981–1983)

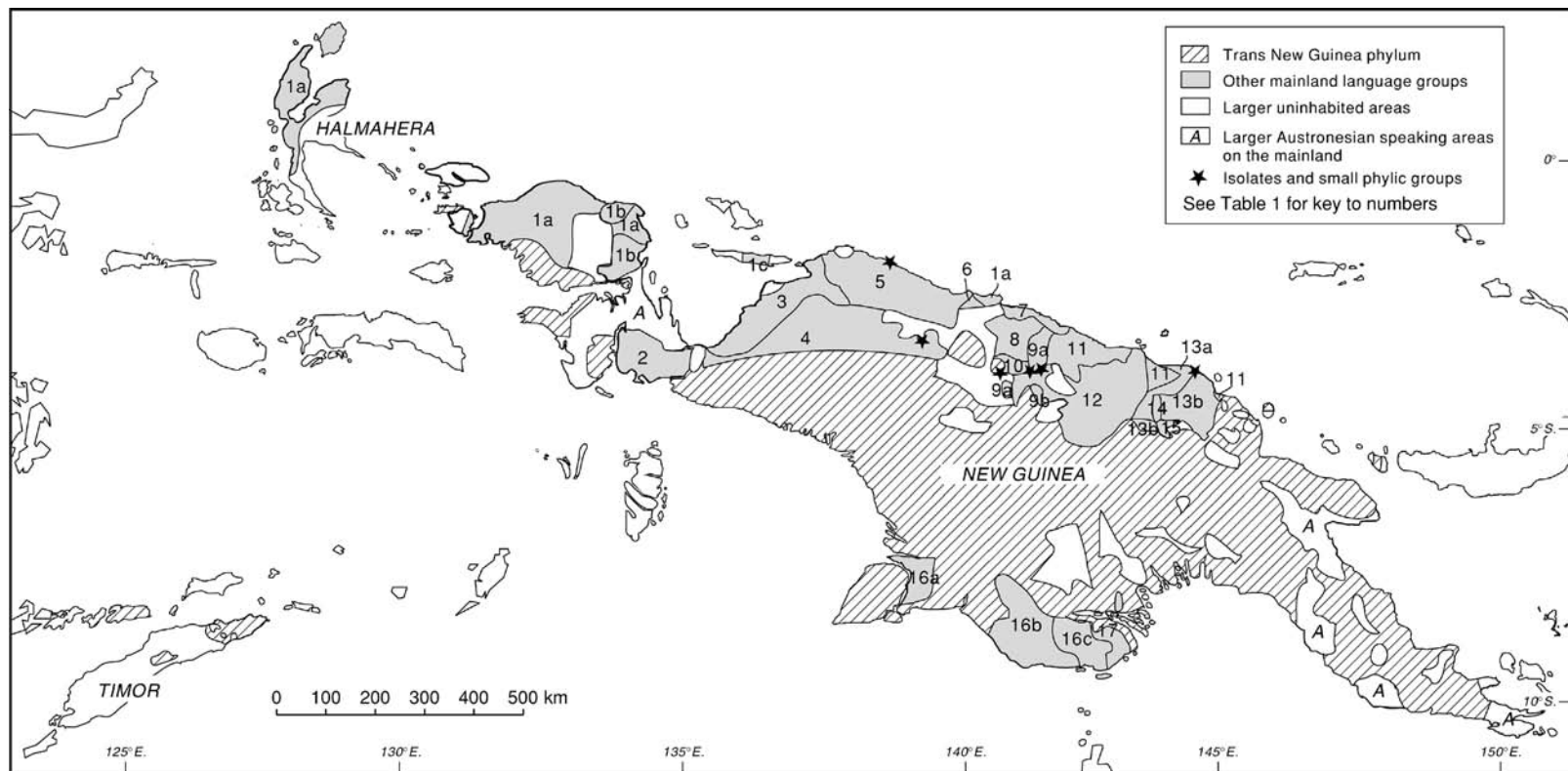


Figure 1 Distribution of Papuan language families in New Guinea and the Timor–Maluku region. Reproduced from Pawley A & Ross M (eds.) *Papuan languages and the Trans New Guinea Family*. Canberra: Pacific Linguistics (forthcoming), with permission.

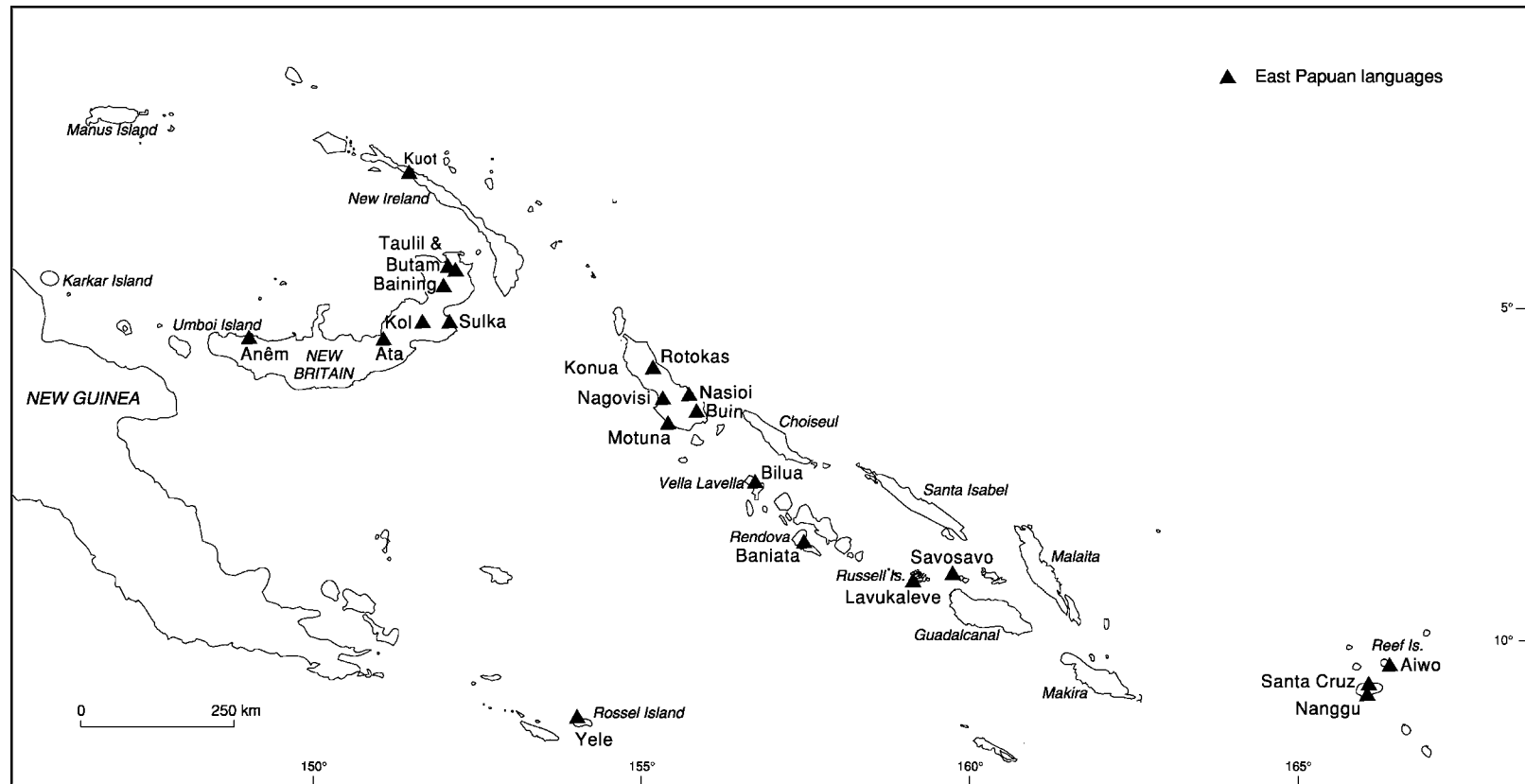


Figure 2 Distribution of Papuan languages in island Melanesia. Reproduced from M Ross, 'Is there an East Papuan phylum?' In A Pawley *et al.* (eds.) (2001). *The boy from Bundaberg. Studies in Melanesian Linguistics in Honour of Tom Dutton*. Canberra: Pacific Linguistics, with permission.

maps in detail the distribution of Papuan languages and language families. However, since this work was compiled several important revisions to the classification have been proposed. This caveat also applies to the information given in *Ethnologue* (Grimes, 2000). The main centers for the study of Papuan languages in recent decades have been the Australian National University's Research School of Pacific and Asian Studies, the University of Sydney, Leiden University, and the Summer Institute of Linguistics's branches in Papua New Guinea and the Indonesian province of Papua (formerly Irian Jaya).

A Short History of Research on Papuan Languages

Until the last decades of the 19th century the languages of the New Guinea area were almost completely unknown to linguists. The imposition of European colonial administrations during that time initiated a period of linguistic research, mainly carried out by missionary scholars. In 1893 the English linguist S. H. Ray observed that some of the languages found in the New Guinea area do not belong to the Austronesian family. Over the next 60 years, as Western exploration of the interior of New Guinea and other large islands proceeded, it became apparent that there were hundreds of such languages and that they were genetically extremely diverse. No families of Papuan languages with more than about 20 members were identified before the 1950s.

Until the end of World War II research on Papuan languages was largely done by scholars with no training in modern linguistics. In the late 1950s a phase of more systematic descriptive and comparative research began. Between 1958 and the 1970s extensive surveys and some in-depth studies of Papuan languages were undertaken by linguists from the Australian National University (ANU). Around 1960 the Dutch linguists Anceaux, Cowan, and Voorhoeve began research in Irian Jaya. Since the Summer Institute of Linguistics established branches in Papua New Guinea in 1956, and in Irian Jaya in 1970, SIL linguists have undertaken descriptive work on some 200 Papuan languages.

This new phase of research yielded a series of preliminary classifications, culminating in a major synthesis by the ANU group (Wurm (ed.), 1975; Wurm, 1982). In 1960 the number of Papuan families was thought to be more than 60. Using mainly lexicostatistical and typological arguments the contributors to Wurm (ed., 1975) reduced the number to 10 'phyla,' along with a number of isolates. (Following the nomenclature often used in lexicostatistical

classifications the ANU group called the highest-order genetic group a 'phylum,' while using 'sub-phylum,' 'stock,' and 'family' to rank subgroups according to percentages of shared cognates.)

In linguistic classification there are lumpers and splitters. Wurm and some of his colleagues can be described as lumpers. The classification in Wurm (ed., 1975), and followed in associated works such as Wurm and Hattori (1981–1983), included three particularly controversial claims. One is that almost 500 Papuan languages can be assigned to a single genetic unit, the Trans New Guinea phylum. If true this would make Trans New Guinea the third largest family in the world in number of members, after Niger-Congo and Austronesian. Second, Wurm (1975) posited an East Papuan phylum consisting of all 20 or so geographically scattered Papuan languages of Island Melanesia plus Yela Dne (Yeletnye) [Yele] of the Lousiade Archipelago, off the southeastern tip of New Guinea. Third, Laycock and Z'graggen (1975) proposed a Sepik–Ramu phylum, to which they assigned almost 100 languages spoken in and around the Sepik–Ramu basin.

Although a good many nonspecialists accepted these proposals uncritically, none was well received by Papuan specialists. All the main reviewers of Wurm (ed., 1975) regarded the Trans New Guinea hypothesis as unproven though not without promise. The Sepik–Ramu hypothesis fell into the same basket. The proposed East Papuan hypothesis was generally viewed as the least plausible of the three.

The most extreme lumper in the Papuan field has been the American linguist, Joseph Greenberg. In a paper drafted much earlier but not published until 1971 Greenberg suggested that all the Papuan languages belong to a vast 'Indo-Pacific' group, to which he also assigned the Andaman Islands and Tasmanian languages. The languages of mainland Australia were excluded. Greenberg's Indo-Pacific proposal rested mainly on a flimsy chain of resemblances in lexical forms (84 sets) and grammatical forms (10 sets). The resemblances are flimsy because the resemblant forms are distributed very unevenly across language groups and because of the lack of means to distinguish shared retentions from chance resemblances and borrowings – one can find a chain of chance resemblances linking any set of sizeable language families. Greenberg divided the Papuan languages of New Guinea into seven major groups, some of which had merit. For example, his 'Central' group resembles the Trans New Guinea (TNG) family in that he assigned to it all the central highlands languages from the Baliem Valley in Irian Jaya to the Huon Peninsula group in Morobe Province, Papua New Guinea. However, evidence for such a group was not given

except as part of the mass of etymologies adduced in support of Indo-Pacific as a whole.

Greenberg's Indo-Pacific proposal drew almost no response from Papuanists. This lack of response, no doubt, reflects (a) extreme skepticism, and (b) the difficulty of disproving a claim of this kind until linguists have established a core of well-defined genetic groups among the languages concerned and have worked out the essentials of their historical development. The main message in the critical reviews of Wurm (ed., 1975) was along the lines of (b). Foley (1986: 3, 213) argued that a properly cautious view should recognize some 60 separate Papuan families which have not been convincingly shown to be related.

Recent Work on the Classification of Papuan Languages

Recently Malcolm Ross compared pronoun paradigms in 605 Papuan languages as a basis for recognizing language families (Ross, forthcoming, 2001, in press). For each family he sought to determine a sequence of innovations in pronoun forms and categories that would yield subgroups. The limitation of Ross's classification is that it relies heavily on a very restricted set of diagnostic criteria. Its strength is that pronoun paradigms have proved to be the most reliable single diagnostic. Ross identifies some 23 to 25 language families and 9 or 10 isolates. The pronominal evidence indicates that the Papuan languages show more genetic diversity than was recognized by Wurm (ed., 1975) but less than was proposed by Foley (1986).

The classification of Papuan groups in **Figure 1** relies heavily on Ross's work but also draws on several other recent studies including Dunn *et al.* (2002), Foley (1986, in press), Pawley (1998, 2001, in press), Reesink (in press), and Terrill (2002).

The Trans New Guinea Family

A slightly reduced version of the TNG family as proposed in the 1970s has been strongly supported by recent work using sounder methods (Pawley, 1998, 2001, in press; Ross, forthcoming) (*see Trans New Guinea Languages*). The main evidence for TNG consists of (a) some 200 putative cognate sets, nearly all denoting so-called 'basic vocabulary,' (b) a body of regular sound correspondences in a sample of daughter languages, which has allowed a good part of the Proto TNG sound system to be reconstructed, (c) systematic form-meaning correspondences in the personal pronouns, permitting reconstruction of virtually a complete paradigm, and (d) widespread resemblances in fragments of certain other grammatical

paradigms. The TNG family, as redefined, contains about 400 languages. Branches of the family occupy the central cordillera that runs the length of New Guinea as far west as the neck of the Bird's Head. They also cover large parts of the southern and, to a lesser extent, the northern lowlands of New Guinea, and have outliers in East Timor, Alor, and Pantar.

Families Confined to North New Guinea

A spectacular degree of linguistic diversity, unmatched anywhere else in the world, is found in north New Guinea between the Bird's Head in Irian Jaya and Madang Province in Papua New Guinea. No fewer than 15 different families plus several isolates are present. The putative Sepik–Ramu family is not supported by Ross's study of pronominal paradigms or by Foley's analysis of a wider range of evidence (Ross, forthcoming; Foley, in press). Foley deconstructs Sepik–Ramu into three unrelated groups: Sepik, Lower Sepik–Ramu, and Yuat. He argues that the Sepik family, containing nearly 50 languages, has its greatest diversity and therefore its original dispersal center in the reaches of the Sepik River above Ambunti. Ross (2000, forthcoming) also recognizes the Sepik and Yuat groups but divides Lower Sepik–Ramu into two possibly unrelated groups: Lower Sepik and Ramu, as well treating Taiap as an isolate. However, he accepts Foley's argument that there are fragments of morphological evidence for uniting Lower Sepik and Ramu. Ross concludes that the distribution of the Ramu and Lower Sepik languages indicates that their diversification predated the regression of the Sepik inland sea some 5000 years ago. As the silt from the Sepik delta filled up this sea Lower Sepik speakers progressively followed the river to the coast.

The Torricelli family proposed by Laycock (1975) is supported. It consists of close to 50 languages, most of which occupy a continuous area of the Torricelli and Prince Alexander Ranges between the Sepik River and the north coast. Languages of the Ndu branch of the Sepik family have expanded north from around the Chambri Lakes and driven a wedge into the Torricelli family, isolating a number of Torricelli languages to the west and south of the Murik Lakes. A small enclave of Torricelli languages also exists on the coast in western Madang Province, isolated from its relatives by a wedge of Ramu languages.

A number of smaller families, each with fewer than 20 languages, have been identified in north New Guinea. These include Skou (spoken on the north coast around the Papua New Guinea-Irian Jaya border), Kwomtari (northwest part of Sandaun [formerly

West Sepik] Province), Left May (situated south of the Kwomtari group around the May River, a tributary of the Sepik), and Amto-Musian (between Kwomtari and Left May). There is some evidence for a Kwomtari-Left May group. Geelvink Bay languages are spoken on the coast of Cenderawasih (formerly Geelvink) Bay. East Bird's Head languages are spoken on the eastern side of the Bird's Head. The West Papuan family, comprising about 24 languages, is represented on the northern part of the Bird's Head at the western end of New Guinea, on Yapen, and on the northern two thirds of Halmahera. There is slight evidence for linking West Papuan and East Bird's Head. On the central south coast of New Guinea at least two groups do not, on present evidence, belong to TNG. Ross refers to these as the South Central family and the Eastern Trans-Fly family.

Island Melanesia

Ross's pronoun study gives no support for Wurm's East Papuan phylum. Instead he finds eight distinct genetic units, including five families, which show a few noteworthy typological similarities, such as a masculine/feminine distinction in 3rd person pronouns (Ross, 2001; Terrill, 2002; Dunn *et al.*, 2002; Wurm, 1982). The Papuan languages of New Britain are divided into an East New Britain family (the Baining dialect chain, arguably more than one language, together with Taulil and Butam), a West New Britain family (Anem and Ata) and two isolates, Sulka and Kol. Another isolate, Kuot, is the only surviving Papuan language in New Ireland, although some neighboring Austronesian languages show what seems to be a Kuot-like substratum. The Papuan languages spoken in Bougainville fall into two families, North Bougainville (Kunua [Rapoi], Kiriaka, Rotokas, and Eivo) and South Bougainville (Nasioi, Nagovisi, Motuna [Siwai], and Buin). On the basis of pronominal resemblances Ross recognizes a Central Solomons family, made up of four languages (Bilua, Baniata, Lavukaleve, Savosavo) scattered across several islands in the main Solomons group. However, there is little else to support such a grouping. In the Santa Cruz group, in the eastern Solomons, there are three languages whose status as Austronesian or Papuan has long been disputed.

Structural Characteristics of Papuan Languages

Good grammars and dictionaries exist for languages in several of the Papuan families. Some representative grammars are Farr (1999) for Trans New Guinea,

Bruce (1984) for the Sepik family, Foley (1991) for the Lower Sepik family, Dol (1999) for the Bird's Head family, van Staden (2000) for the West Papuan family, Onishi (1995) for Motuna of the South Bougainville family, and Terrill (2003) for the Central Solomons family.

Because of their genetic diversity it is hard to generalize about the structure of Papuan languages. However, in New Guinea there are many diffusion areas where certain structural features as well as lexicon have spread across language family boundaries.

The phonemic inventories of Papuan languages range from among the smallest in the world (Lakes Plains of Irian Jaya and Rotokas of Bougainville each has 11 segmental phonemes) to quite large (Yela Dne of Rossel). A five vowel system /i, e, a, o, u/ is the commonest, although a number of languages have various types of six, seven, and eight vowel systems. Word-tone or pitch-accent contrasts are fairly common among Papuan languages, for example in the TNG, Lake Plains, West Papuan, Geelvink Bay, and Skou families (Donohue, 1997).

In most Papuan families the preferred order of constituents in verbal clauses is SOV. Notable exceptions are the Torricelli family, East Bird's Head, some members of the West Papuan family spoken in Halmahera, and three of the languages of the Central Solomons group, where SVO order is usual. The Halmahera and Central Solomons languages with SVO order have been strongly influenced by Austronesian neighbors.

In most Papuan families grammatical relations like subject, object, and location are signaled by adpositions or word order, or the presence on the verb of person-number affixes for subject and object. Most languages organize pronominal affixes to show a nominative-accusative (or dative) contrast. Only a few languages have a true ergative-absolutive alignment for verb pronominals. Some TNG languages optionally mark a wilful or focused agent by what is otherwise the instrument postposition.

Pronominal systems vary considerably across and even within families and there is often a discrepancy between the kinds of distinctions made in independent pronouns and in verbal affixes. TNG languages typically distinguish roots for 1st, 2nd, and 3rd person, adding number markers for plural (some languages also distinguish a dual and, less commonly, a paucal). An exclusive/inclusive contrast is absent from most Papuan families. It is restricted to groups such as West Papuan, certain Torricelli languages, and a few isolates. In at least some cases this contrast may be a feature borrowed from Austronesian neighbors.

Almost all Papuan families distinguish sharply between noun roots and verb roots. Generally a verb root cannot be used as a noun without derivational morphology and vice versa. In certain TNG languages verb roots are a small closed class, with somewhere between 60 and 150 members. The densest concentration of these languages is in the Chimbu–Wahgi and Kalam–Kobon subgroup, located in the Central Highlands and the contiguous Schrader Ranges. Most TNG languages and some other Papuan families augment their stock of verbs by forming complex predicates consisting of a verbal adjunct or coverb plus a verb root. Verbal adjuncts are uninflected bases that occur only in partnership with a verb, often being restricted to one or a few verbs. Most TNG and Sepik and Lower Sepik–Ramu languages also make heavy use of serial verb constructions consisting of consecutive bare verb roots.

Verb morphology is typically of medium to extreme complexity. Most languages carry suffixes marking tense, aspect and mood, and person–number of subject, and some also carry prefixes marking the object. But there are exceptions: agreement affixes are lacking in Lakes Plain and Lower Ramu languages, in the TNG languages of the Timor region, and in certain Geelvink Bay languages. In TNG languages there is often a degree of fusion of the subject marking and TAM suffixes. A degree of morphological complexity is found in some languages of the Sepik–Ramu basin, such as Yimas (Foley, 1991), Alamlak (Bruce, 1984), Barupu of the Skou family, and in the Kainantu subgroup of TNG, all of which show polysynthetic characteristics.

A prominent feature of most TNG languages is the marking of ‘medial’ verbs for switch reference and relative tense. Whereas sentence-final verbs head the final clause in a sentence and carry suffixes marking absolute tense-aspect-mood and person–number of subject, medial verbs head nonfinal coordinate-dependent clauses and carry suffixes marking (a) whether the event denoted by the medial verb occurs prior to or simultaneous with that of the final verb, and (b) ‘switch-reference,’ i.e., whether that verb has the same subject or topic as the next clause.

In the Torricelli and Lower Sepik–Ramu families and in certain other small groups of north-central New Guinea, nouns carry complex inflections, marking number distinctions and noun classes. Noun class systems are an areal feature of languages belonging to diverse families in the Sepik–Ramu basin in New Guinea. The Torricelli, Sepik, and Lower Sepik–Ramu families have upwards of 10 noun classes. Most classes are assigned phonologically, according to the final segment of the stem. The isolate Burmeso, in northern Irian Jaya, has six genders and six noun

classes, marked simultaneously. A few TNG languages that are neighbors of members of the Lower Sepik–Ramu group have acquired noun classes. Noun classes are also found in Bougainville and in the central Solomons. Gender classes, usually just masculine versus feminine, are distinguished in nouns in West Papuan and Skou (shown by agreement prefixes) and in the Sepik family and a small minority of Trans New Guinea languages (marked by concord suffixes). Feminine is usually the unmarked gender.

Some Trans New Guinea languages use existential verbs like ‘stand,’ ‘sit,’ ‘lie,’ and sometimes others like ‘hang,’ ‘carry,’ and ‘come’ as quasi-classifiers of nouns. Nouns select a verb according to their shape, posture, size, and composition. However, the classification is not absolute for the noun but has some flexibility relative to the situation of the referent. Papuan languages show a wide variety of numeral systems, including the ‘Australian’ system (1, 2, 2+ 1, 2+ 2), quaternary, quinary, vigesimal, and various kinds of body-part systems.

Explaining the Diversity and Distribution of the Papuan Languages

Why is the New Guinea area so linguistically diverse, in terms both of the number of apparently unrelated genetic stocks and the number of individual languages? One major factor is the very great time depth available for *in situ* diversification. Archaeology has shown that humans reached New Guinea and Australia (then joined as a single continent, Sahul) upwards of 40 000 years ago (Spriggs, 1997). By 40 000 to 36 000 years ago people crossed from New Guinea to New Britain, the nearest part of Island Melanesia, and from New Britain to New Ireland. By 29 000 to 28 000 years ago people had made the 180 km crossing from New Ireland to the northern end of Bougainville. The initial phase of human expansion into the southwest Pacific got no further than the main Solomons chain, which ends at Makira (San Cristobal). There is no evidence that humans settled any part of Remote Oceania, i.e., the Pacific Islands beyond the main Solomons chain, until about 3200 years ago.

A second force aiding diversification resides in social and political organization. In the New Guinea area political units were small, probably seldom larger than a collection of kinship groups or one or two villages containing a few hundred people. No unit had the political and economic power to dominate a large area. Neighboring polities were often hostile. A third factor is geographic barriers. In New Guinea, New Britain, and Bougainville, in particular, heavily forested mountain ranges and extensive

swamps imposed natural limits to communication. Substantial ocean gaps between islands provided natural points of linguistic fission for people who lacked efficient ocean-going craft.

A fourth factor, which kept established language families from being overrun by invading groups, is the lengthy isolation of much of the New Guinea area itself. The evidence of archaeology and population genetics (Friedlaender *et al.*, in press) indicates that the people of New Britain, New Ireland, and Bougainville had little contact with the rest of the world for tens of millennia following initial settlement. The same may have been true, though to a lesser extent, of populations inhabiting the interior of New Guinea. One can speculate that some of the diverse language stocks of both the New Guinea mainland and Island Melanesia continue the languages of the earliest, late Pleistocene settlers in these regions. As Australia and New Guinea were connected as recently as about 8000 years ago one might expect to find traces of old connections with Australian languages, but no solid evidence has been found (see Foley, 1986 for some speculations).

Two major expansions show up in the linguistic record for the New Guinea area. The TNG family is exceptional among Papuan families in its large membership and wide geographic spread. The great diversity among its subgroups shows that TNG is a very ancient family which, according to glottochronological estimates (admittedly not very reliable) began to diverge some 8 to 12 millennia ago. The distribution of subgroups suggests that its most likely primary center of diversification is the central highlands of Papua New Guinea. It seems unlikely that the TNG family would have achieved its present remarkable distribution unless its speakers possessed cultural advantages that enabled them to pioneer permanent settlement of the heavily forested high valleys of the central cordillera. The key advantage may have been agriculture. Archaeological work indicates the presence of full-scale agriculture near Mt. Hagen in the Upper Wahgi Valley, probably by 10 000 years ago and certainly by 7000 years ago (Denham *et al.*, 2003).

However, it is striking that speakers of TNG languages made few inroads into the Sepik provinces and the western half of Madang province and the Bird's Head. These areas are dominated by other, much smaller families. It is reasonable to suppose that at the time of the TNG expansion these regions were already inhabited by speakers of some of the non-TNG languages that are still represented there.

A second major linguistic expansion occurred in the 2nd millennium B.C. when Austronesian speakers arrived in the New Guinea area. This event shows up

Table 1 Papuan families identified in **Figures 1 and 2**

| | |
|-----|---|
| 1. | 'extended West Papuan' (?) (1a) West Papuan (1b) East Bird's Head, Sentani (1c) Yawa |
| 2. | Mairasi |
| 3. | Geelvink Bay |
| 4. | Lakes Plain |
| 5. | Orya-Mawes-Tor-Kwerba |
| 6. | Nimboran |
| 7. | Skou |
| 8. | Border |
| 9. | Left May-Kwomtari (9a) Kwomtari (9b) Left May |
| 10. | Senagi |
| 11. | Torricelli (three separate areas) |
| 12. | Sepik |
| 13. | Ramu-Lower Sepik (13a) Lower Sepik (13b) Ramu |
| 14. | Yuat |
| 15. | Piawi |
| 16. | South-Central Papuan |
| 16a | Yelmek-Maklew |
| 16b | Morehead-Upper Maro |
| 16c | Pahoturi |
| 17. | Eastern Trans-Fly |
| 18. | Trans New Guinea |
| 19. | Yela Dne-West New Britain (?) |
| 20. | East New Britain |
| 21. | North Bougainville |
| 22. | South Bougainville |
| 23. | Central Solomons |

clearly in the archaeological record (Spriggs, 1997). There is good reason to think that 3000 years ago northern Island Melanesia contained many more Papuan languages than it does now. Whereas this region now harbors about 150 Austronesian languages (all belonging to the large Oceanic subgroup) only about 21 Papuan languages survive there. None are present in the Admiralty Islands and only one in New Ireland. Although they came to dominate the smaller islands of Melanesia, Austronesian languages had much less impact in New Guinea. There they are mainly confined to offshore islands and to certain patches along the north coast and in southeast Papua.

There are abundant signs that the Austronesians at first had a similar, marginal distribution in Island Melanesia. However, the eventual outcome was very different. In due course the Admiralty Islands and most of New Britain, New Ireland, Bougainville, and the Solomons became Austronesian-speaking, though not without a good deal of linguistic and cultural exchange between immigrants and aboriginal populations (Dutton and Tryon, 1994).

In much of Island Melanesia it seems that the interaction between Austronesian and Papuan speakers was of a kind that led to widespread language shift. With few exceptions the shifts appear to have been cases of communities that formerly spoke Papuan languages adopting Austronesian languages while maintaining much of their biological and social distinctiveness. As to the mechanisms of language shift, there have as yet been few studies.

Bibliography

- Bruce L (1984). *The Alamlak language of Papua New Guinea (East Sepik)*. Canberra: Pacific Linguistics.
- Carrington L (1996). *A linguistic bibliography of New Guinea*. Canberra: Pacific Linguistics.
- Denham T P *et al.* (2003). 'Origins of agriculture at Kuk Swamp in the highlands of New Guinea.' *Science* 201, 189–193.
- Dol P (1999). A grammar of Maybrat, a language of the Bird's Head, Irian Jaya, Indonesia. Ph.D. thesis, University of Leiden.
- Donohue M (1997). 'Tone systems in New Guinea.' *Linguistic Typology* 1, 347–386.
- Dunn M, Reesink G & Terrill A (2002). 'The East Papuan languages: a preliminary typological appraisal.' *Oceanic Linguistics* 41, 28–62.
- Dutton T E & Tryon D T (eds.) (1994). *Language contact and change in the Austronesian world*. Berlin: Mouton de Gruyter.
- Farr C (1999). *The interface between syntax and discourse in Korafe: a Papuan language of Papua New Guinea*. Canberra: Pacific Linguistics.
- Foley W (1986). *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley W (1991). *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- Foley W (2000). 'The languages of New Guinea.' *Annual Review of Anthropology* 29, 357–404.
- Foley W (2005). 'Linguistic prehistory in the Sepik–Ramu Basin.' In Pawley *et al.* *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics.
- Friedlaender J *et al.* (2005). 'Mitochondrial genetic diversity and its determinants in Island Melanesia.' In Pawley *et al.* *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics.
- Greenberg J H (1971). 'The Indo-Pacific hypothesis.' In Sebeok (ed.). 807–871.
- Grimes B (ed.) (2000). *Ethnologue: languages of the world* (14th edn.) (2 vols). Dallas: SIL International.
- Laycock D (1975). 'The Torricelli phylum.' In Wurm (ed.). 767–780.
- Laycock D & Voorhoeve C (1971). 'History of research in Papuan languages.' In Sebeok (ed.). 509–540.
- Laycock D & Z'graggen J (1975). 'The Sepik–Ramu phylum.' In Wurm (ed.). 731–763.
- Onishi M (1995). A grammar of Motuna (Bougainville, Papua New Guinea). Ph.D. diss., Australian National University.
- Pawley A (1998). 'The Trans New Guinea phylum hypothesis: a reassessment.' In Miedema J, Ode C & Dam R A C (eds.) *Perspectives on the Bird's Head of Irian Jaya, Indonesia*. Amsterdam: Editions Rodopi. 655–689.
- Pawley A (2001). 'The Proto Trans New Guinea obstruents: arguments from top-down reconstruction.' In Pawley, Ross & Tryon (eds.). 261–300.
- Pawley A (forthcoming). 'The chequered career of the Trans New Guinea hypothesis: recent research and its implications.' In Pawley, Attenborough *et al.*
- Pawley A, Ross M & Tryon D (eds.) (2001). *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*. Canberra: Pacific Linguistics.
- Pawley A, Attenborough R, Golson J & Hide R (in press). *Papuan pasts: studies in the cultural, linguistic and biological history of the Papuan-speaking peoples*. Canberra: Pacific Linguistics.
- Reesink G (2005). 'West Papuan languages: roots and development.' In Pawley *et al.* *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics.
- Ross M (forthcoming). 'Pronouns as preliminary evidence for grouping Papuan languages.' In Pawley A, Ross M & Osmond M (eds.) *Papuan languages and the Trans New Guinea family*. Canberra: Pacific Linguistics.
- Ross M (2001). 'Is there an East Papuan phylum? Evidence from pronouns.' In Pawley, Ross & Tryon (eds.). 301–321.
- Ross M (2005). 'Pronouns as a preliminary diagnostic for grouping Papuan languages.' In Pawley, Attenborough *et al.* *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics.
- Sebeok T E (ed.) (1971). *Current Trends in Linguistics 8: Linguistics in Oceania*. The Hague: Mouton.
- Spriggs M (1997). *The Island Melanesians*. Blackwell: Oxford and Cambridge, MA.
- Terrill A (2002). 'Systems of nominal classification in East Papuan languages.' *Oceanic Linguistics* 41, 63–88.
- Terrill A (2003). *A grammar of Lavukaleve*. Berlin: Mouton de Gruyter.
- van Staden M (2000). Tidore: a linguistic description of a language of the North Moluccas. Ph.D. diss., University of Leiden.
- Wurm S (1975). 'The East Papuan phylum in general.' In Wurm (ed.). 783–804.
- Wurm S (1982). *Papuan languages of Oceania*. Tubingen: Gunter Narr.
- Wurm S (ed.) (1975). *New Guinea area languages* (vol. 1). *Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics.
- Wurm S & Hattori S (eds.) (1981–1983). *Language atlas of the Pacific area* (2 vols.). Canberra: Australian Academy for the Humanities in collaboration with the Japanese Academy.

Pashto

D Septfonds, National Institute of Oriental Languages,
Paris, France

© 2006 Elsevier Ltd. All rights reserved.

There is also a large diaspora in the Gulf countries, particularly in Dubai, and in Europe, the United States, and Australia.

Origin and History

Pashto is spoken by some 40 million people living on both sides of the border between Pakistan and Afghanistan, the famous Durand line, which has given rise to many conflicts. This line was drawn in 1893 following an agreement between Afghanistan and British India, which determined the southern limits of Afghanistan and divided Pashtun territory between Afghanistan and British India – Pakistan since Partition in 1947.

Pashto is the language of the tribes that founded the Afghan state in 1747: the Pashtuns or, according to the term that prevailed in British India, the Pathans (Indianized form of the plural of Pashtuns).

Pashto is the main language spoken in Afghanistan and one of the two official languages of the country, the other being Dari or Afghan Persian. Pashto, which is mainly spoken south of the mountain range of the Hindu Kush, is reportedly the mother tongue of 60% of the Afghan population. Many Pashto-speaking pockets are also found in the north and the northwest of the country where Pashtuns were transferred in the late 19th century and given land.

In Pakistan, Pashto, which is spoken by 20–25 million people, has the status of a regional language. While the majority of the Pashtuns live in the North-West Frontier Province (NWFP, capital Peshawar), in Baluchistan (capital Quetta), or in the Federally Administered Tribal Areas (FATA) – the Pashtun area being roughly at the East of the Indus – Karachi, where about two million people speak Pashto, remains the main Pashtun metropolis.

Dialectology

From a strictly genetic point of view, Pashto, an Indo-European language, belongs to the northeastern group of Iranian languages. From one dialect to another, its morphosyntactic structure does not show any major variation. Their classification is based on phonological criteria and depends on the pronunciation of **بن** /*ṣ̌*/ and **ښ** /*ḡ*/ letters. These consonants are pronounced differently according to the regions. This constitutes a first isoglossic line: the most visible and the most notable, since it can be observed in the script. In the A zone (eastern /*maf̌reqi*/), they are pronounced respectively /*g*/ and /*x*/. ‘woman’ is pronounced /*xája*/, ‘beard’ is pronounced /*girá*/: these dialects are known as ‘hard.’ This is found in the English transcription ‘Pukhtu’ (*kh* = *x* = **بن**). In the C zone (western /*maṽrebi*/), they are pronounced /*z*/ and /*ʃ*/, and sometimes reduced to /*j*/ and /*ʒ*/ (Ghazni). ‘woman’ is pronounced /*ʃája*/, beard is pronounced /*zirá*/: these dialects are known as ‘soft’ dialects or ‘Pushtu’ (*sh* = *ʃ*, *š* = **بن**). Both these dialects are written in the same script and the speaker is free to read his own way, with his own ‘accent.’ This unity of script allows the definition of a standard Pashto consisting of A- and C-type dialects, whether ‘hard’ or ‘soft,’ from Kandahar or from Peshawar (Table 1).

On the other hand, crossing the line separating ‘soft’ Pashto from ‘hard’ Pashto, another isogloss exists that defines a B zone known as intermediary or central (/*mandʒanəy*/). This zone, which does not present such clear unity as the zones mentioned above

Table 1 Dialects

| | | Zone C | | Zone B | | Zone A | |
|--------------------|-------------|--------------------------|---------------------|----------------------------|------------------------|---------------------------|--|
| <i>ṣ̌</i> | بن | <i>ʃ</i> | <i>j</i> | <i>ṣ̌</i> = [ç] | <i>x</i> | | |
| <i>ḡ</i> | ښ | <i>z</i> | <i>ʒ</i> | <i>ḡ</i> = [j] | <i>g</i> | | |
| Standard Pashto | Afgh Pak | Kandahar Quetta | Ghazni | | Djalálábád Peshawar | | ‘father’ / <i>pla:rl</i> ‘mother’ / <i>morl</i> ‘daughter’ / <i>lurl</i> |
| Nonstandard Pashto | Afgh Pak | | Waziristan Bannu | Paktyā | | | ‘father’ / <i>plorl</i> ‘mother’ / <i>merl</i> ‘daughter’ / <i>lirl</i> |
| | | Zone C <i>maṽrebi</i> | | Zone B <i>mandʒanəy</i> | | Zone A <i>maf̌reqi</i> | |

Afgh = Afghanistan/Pak = Pakistan.

as far as the pronunciation of the consonants is concerned (نښ /x/ or /f/ and ځ /ǵ/ or /ʒ/), nevertheless clearly contrasts with them due to a very particular pronunciation of the vowels of standard Pashto. It is the Waziri metaphony, taking its name from the Wazir tribes among whom it is well attested: pronunciation /o/ for /a:/, /e/ for /o/, /i/ for /u/. If this pronunciation had to be written, it would impose a script contrary to the entire orthographic tradition. This type of Pashto is not written. As a consequence, speakers belonging to this zone use a different type of Pashto when they have to communicate with other Pashtuns (from zones A or C). We find here a perfect example of diglossia: they use a variant of their language that is better recognized because it is written; better recognized, though not more prestigious because these dialects have a strong value as indicators of identity.

Another residual variant of Pashto – Wanetsi – is spoken in Pakistani Baluchistan. This archaic idiom, which has hardly been described, is virtually unintelligible to other Pashtuns.

Pashto Script and Orthography

Pashto literature dates from the 16th century. The publication in 1975 of a facsimile of a manuscript, supposedly dating from 1886, places the beginning of Pashto literature as far back as the 8th century. The authenticity of this poetry anthology – The hidden treasure /pəʃa xaza:na/ – is much debated. Pashtun land was then divided between the Safavid and the Moghul empires. The literary model was Persian, and Pashto scholarly literature has inherited Arabo-Persian poetical genres and meters. This literature starts with Khayr ul Bayan (The best discourse) of Bayazid Ansari; the most ancient manuscript dates

back to 1651. Bayazid Ansari, known as Pir Roshan (The luminous master, 1524–1579), the founder of a politico-religious movement considered a heresy, waged war on Delhi.

From the point of view of development of the script, this represents the birth of the first tradition. Three other subsequent traditions can be distinguished, with some overlap and parallel developments. The first is the tradition of Khushal Khan Khattak (1613–1689) – the poet warrior, father of Pashto literature – and of his descendants, which constitutes in itself a literary tradition. A standard tradition followed in the 19th and 20th centuries, mainly in Pakistan, with some characteristic features of the Urdu script. Finally, the tradition of ‘modern’ script has developed since 1936 in Afghanistan.

Nowadays, on both sides of the border, the orthographic standard is the Afghan scholarly standard, which has drawn on the Persian script since the early 1990s.

In all these cases, the script is the Arabic script adapted to the needs of a language that has phonemes unknown in Arabic: these phonemes are common to Pashto, Persian, and Urdu (پ /p/, ټ /tʃ/, ژ /ʒ/, ګ /g/), to Pashto and Urdu (ټ /t/, ډ /d/, ځ /ǵ/), while some letters are particular to Pashto (نښ /x/, ځ /ǵ/, څ /ts/, ځ /dz/, ځ /ɲ/). (See phonemes in bold in Table 2.)

Basic Phonology

Pashto is a language with free accentuation.

A remarkable feature of Pashto is a series of retroflex consonants (/ʈ/, /ɖ/, /ʂ/, /ʐ/, /ɳ/), which is exceptional in Iranian languages; Pashto also has many word-initial clusters that cannot exist in Persian Table 3.

Table 2 Consonants

| | <i>Bilabial</i> | <i>Dental/Alveolar</i> | <i>Retroflex</i> | <i>Velar</i> | <i>Uvular</i> |
|------------|-----------------|------------------------|------------------------------------|--------------|----------------|
| Plosive | p b | t d | ʈ ɖ | k g | q ^b |
| Affricate | | ts dz | | | |
| | | tʃ ɖʒ | | | |
| Fricative | f | s z | ʃ^a ʒ^a | x ɣ | h |
| | | | ʃ ʒ | | |
| Nasal | m | n | ɳ | | |
| Liquid | | | ɭ | | |
| | | r | | | |
| Semivowels | w | | y | | |

^aCf. ‘Dialectology.’

^bIn italics: ‘the elegant phonemes.’

These phonemes are not native pashto sounds. They occur in the speech of educated speakers only (in Arabic and Persian loan words). /q/ varies with /k/ in a stylistically determined alternation, /f/ with /p/ and /h/ – lengthening a preceding vowel /a/ – with zero.

Basic Morphology

Nouns

Nouns in Pashto are inflected for gender (masculine, feminine), number (singular, plural) and case (direct = nominative, oblique, vocative). ‘Prepositions’ (preposition, circumposition, and postposition) govern the oblique case.

Pronouns

Pronouns are inflected according to Table 4. There are three series: personal pronouns (tonics); personal clitics (used as ‘actant’ – subject or object – and also in possessive constructions); and verbal inflections.

These forms are divided into weak and strong: /ə/ vs. /zə/; /-me/ vs. /ma:/ . A particular weak series, the series of directional pronouns, corresponds to the strong series /ma:/, when the latter is governed by ‘prepositions.’

Pashto Verbs

Pashto verbs have two stems, one for present tense forms and one for past tense forms. The infinitive is derived from the past stem by adding /əl/. It is a masculine plural, for instance, ‘to see’ *lid-l* /win/ (past stem /lid/, infinitive /lid-l/, present stem /win/). Verbs are inflected for person, number and gender (cf. Table 4).

Table 3 Vowels

| Front | Central | Back |
|-------|---------|------|
| i | | u |
| e | ə | o |
| a | | a: |

Table 4 Personal markers in STD Pashto

| | NP | | Enclitic | Directional | Personal ending | |
|-----------|-----------------|---------------|---------------|-------------|-----------------|--------|
| | NOM | OBL | (OBL) | | Present | Past |
| 1 sing | zə | ma: | -me | ra: | | əm(a) |
| 2 sing | tə | ta: | -de | dar | | e |
| 3MASC | day | də | | | | ə/ay/ə |
| 3 sing | -da: | de | -ye | war | -i | |
| 3FEMSing | | | | | | a |
| 1 pl | | munğ | -mo | ra: | u | |
| 2 pl | | ta:se / ta:so | -mo | dar | əy | |
| 3.MASC.PL | | | | | | ə |
| 3.PL | | duy | -ye | war | i | |
| 3 FEM.PL | | | | | | e |
| | Strong pronouns | | Weak pronouns | | | |

From the present stem, two presents are formed (imperfective vs. perfective) and two imperatives (perfective vs. imperfective). From the past stem, two pasts are formed (perfective vs. imperfective). This aspectual perfective vs. imperfective opposition is dominant and is found in the entire verbal system.

In addition to these simple tenses, there are three processes of ‘auxiliation’: the perfect system, the capacitive system (or ‘potential,’ which expresses the capacity; the verb ‘can’ does not exist in Pashto); and the passive.

The system is enriched by the combination of these basic forms with several modal-aspectual enclitic particles (eventual /ba/, injunctive /de/, assertive /xo/), which as such are placed in second position in the utterance (cf. ‘Basic Syntax’ below).

There are two forms of conjugation in Pashto: one for simple verbs (for example ‘to see’ *lidl*) and one for denominative verbs: ‘white’ /*spin*/ gives ‘to whiten’ /*spinedl*/ (intransitive) and ‘to whiten’ /*spinawl*/ (transitive); ‘in good shape’ /*dʒoɽ*/ gives ‘to build up, to get better’ /*dʒoɽedl*/, ‘to heal, to build’ /*dʒoɽawl*/ . There are no more than 150 simple verbs: however, the list of compound verbs is open and productive. To these are added many verbal phrases, mainly with the verb ‘to make’ /*kawl*/ (‘to play’ = ‘game make’ /*lobe kawl*/, ‘to sleep’ = ‘sleep make’ /*xob kawl*/ vs. ‘to dream’ = ‘sleep see’ /*xoblidl*/).

Basic Syntax and Typology

Order of Terms

Within the nominal syntagma, the order of terms is always ‘determinant’ (head modifier) + ‘determined’ (head noun). The process is recursive toward the left. However, two different structures can be distinguished, according to whether the determinant is a noun or an adjective. If the determinant is a noun, it is

Table 5 Nominal determination

| Head modifier | | Head noun | |
|---------------------|-------------|-------------|-----------------------------------|
| də mə:zi'gar | zyeɽ | Imar | (The) yellow sun |
| də mə:zi'gar | zyeɽ | Imar | (The) sun of the afternoon |
| of afternoon.OBL | yellow.NOM | Imar | (The) yellow sun of the afternoon |
| | | SUN.NOM | |

preceded by the preposition /də/ and occurs in the oblique case. If the determinant is an adjective, it directly precedes the determined, without preposition. There is agreement in gender, number, and case (Table 5).

SOV

Pashto is an SOV type language; however, this order is usually breached by the Indo-European rule of the raising of enclitics in second position in the utterance; more exactly, after the first nominal syntagma ('noun phrase'), without any particular syntactic link with the latter. The rule is purely formal: it does not correspond to any semantic pattern.

- (1) ma: da: 'xədəza 'wə-li'dəl-a
 S O Vo
 'me'.OBL this.NOM PERF-
 woman. see.PAST-3FEM.sing
 FEM.NOM
 'It is I who saw this woman'
- (2) da: 'xədəza-me 'wə-li'dəl-a
 O-s Vo
 this.NOM woman. PERF-see.PAST-3
 FEM.NOM- FEM.sing
 ENCL.1.sing
 'I saw this woman'
- (3) 'wə-me-li'dəl-a
 PERF-s-V-o
 PERF-ENCL.1.sing-see.PAST-3FEM.sing
 'I saw her'

Ergativity

Pashto shows a type of 'split ergativity' determined by tense.

In present and past tenses, the subject of the intransitive verb is in the direct case and the verb agrees with it, as in example (4). The nominal term may be missing, as indicated by the brackets.

If this is compared with transitive verbs, it will be seen that in present tenses (formed with the present stem) the construction is accusative and in past tenses (formed with the past stem, both simple and compound forms) the construction is ergative, whatever the aspect.

The construction is accusative in the present because the subject of the transitive verb behaves in the same way as the subject of the intransitive verb. In example (5), it is in the direct (nominative) case and the verb agrees with it.

The construction is ergative in the past because the object of the transitive verb behaves in the same way as the subject of the intransitive verb. Thus, in examples (1)–(3), it is the object term that is in the direct case and the subject term that is in the oblique case. Moreover, the verb agrees with this object, whether it is given, as in (1) and (2), or not (3).

- (4) (zə) j-əm [tl-əm]
 S V s
 'me'.NOM go.PRES-1/sing [go.PAST]
 'It is I who am going [was going]'
- (5) (zə) da: 'xədəza win-əm
 S O V s
 'me'.NOM cette.NOM voir.PRES-1/sing
 femme.FEM.NOM
 'It is I who see this woman'

Antipersonal Verbs

In the past, verbs of this class construct in such a way that the 'subject' is in the oblique and the 'object' is referenced in the verb form by an absolutive marker but cannot be represented by an NP.

- (6) spi 'wə-vəpəl
 dog.OBL PERF-see.PAST-3 PL
 'The dog barked'

The verb contains a marker of 3rd person masculine plural, which refers to nothing.

The form of this construction is clearly ergative, like the biactant one. In the present, this very small group of verbs (e.g., to laugh, to bark, to jump, to cry, to swim, to bathe) has an intransitive construction.

Differential Object Marking

Pashto also possesses a 'differential object marking.' In the present, according to the place of the object in the *nominal hierarchy*, it is placed either in the oblique case (1st and 2nd person, (7)) or in the direct case (from 3rd person to indefinite; (5)).

Table 6 Landey

گل مې په لاس کې مړاوی کېږي^۱

پردي وطن دی زه یې چا ته ونیسمه

gwəl me pə 'la:s ke mʔa:way 'ki:gi.

praday wa'tan day, zə ye 'tʃa: ta wə-ni'səma?

'The flower withers in my hand'

'This is a foreign land, to whom shall I give it?'

| | | |
|-----------------------|---------------------|-----------------------------|
| (7) (zə) | ta: | win-əm |
| S | O | V s |
| 'me'. _{NOM} | you. _{OBL} | see. _{PRES-I/sing} |
| 'It is I who see you' | | |

The Landey

It is impossible to talk about the Pashtun world without mentioning a popular poetical genre: the landey, literally 'short.' Often sung, their rhythm is invariable; every one knows a number of landeys and is able to compose new ones (accented syllables are in bold (Table 6)).

--- 4 --- 8 -

--- 4 --- 8 --- 12 -

Bibliography

- Bellew H W (1867). *A grammar of the Pukhto or Puksho language*. London: (Reprint 1983) Peshawar.
- Biddulph C E (1890). *Afghan poetry of the seventeenth century, being selections from the poems of Khushal Khan Khatak*. London.
- Caroe O (1965). *The Pathans*. London: Macmillan Press.
- Darmesteter J (1888–1890). *Chants populaires des Afghans*. Paris: (Reprint, Société Asiatique, collection des ouvrages orientaux, 2° série, Amsterdam: Philo Press).
- Elfenbein J (1984). 'The Wanetsi Connexion.' *JRAS* 54–76, 229–241.
- Grierson G A (1921). *Linguistic Survey of India, vol. 10*. Calcutta.
- Kabir H & Akbar M (1999). *Dictionnaire pashto-français*. Paris: L'Asiathèque.

- Kieffer C M (1983). 'Afghanistan'. In *Languages, Encyclopedia Iranica* 1/5, 501–516.
- Lorenz M (1979). *Lehrbuch des Pashto (Afghanisch)*. Leipzig.
- Lorimer J G (1902). *Grammar and vocabulary of Waziri Pashto*. Calcutta: Office of the Superintendent of Government Printing, India [Pakistan].
- Mackenzie D N (1959). 'A standard Pashto.' *BSOAS* 231–235.
- Mackenzie D N (1965). *Poems from the Divan of Khushal Khan Khattak*. London: George Allen and Unwin Ltd.
- MacKenzie D N (1997). 'The development of the Pashto script.' In Akiner S & Sims-Williams N (eds.) *Languages and scripts of Central Asian*. London: School of Oriental and African Studies. 137–143.
- Manalaï N (1987). 'La métrique du pashto.' In *Cahiers de Poétique Comparée* 15. Paris: Publications Langues'O. 102–153.
- Manalaï N (2001). "Tresses et labyrinthe (contes pachto)" suivi de "Truands et filles d'ogresses" par D. Septfonds.' Paris: CEREDAF.
- Morgenstierne G (1960). 'Afghân, la langue pashto.' *EI*.
- Morgenstierne G (2003). *A new etymological vocabulary of Pashto*. Wiesbaden: Reichert.
- Penzl H (1955). *A grammar of Pashto: a descriptive study of the dialect of Kandahar, Afghanistan*. Washington: American Council of Learned Societies.
- Raverty H G (1980 [1860]). *A dictionary of the Pukhto, Pushto, or language of the Afghans*. London: William and Norgate. (Reprint: Indus Publications, Karachi.)
- Septfonds D (1989). 'Du Wanetsi au pashto standard: /kawəl/ ou /kʔəl/?' *Etudes irano-aryennes offertes à Gilbert Lazard*. Paris. 297–313.
- Septfonds D (1994). *Le Dzadrâni, un parler pashto du Paktyâ (Afghanistan)*. Louvain-Paris: Peeters.
- Septfonds D (1997). 'Constructions anti-impersonnelles en pashto.' *SILTA XXVI* (2), 271–306.
- Skjærvø O (1989). 'Pashto.' In *Compendium Linguarum Iranicarum*. 384–410.
- Tegey H & Robson B (1996). *A reference grammar of Pashto*. Washington, DC: Center for Applied Linguistics.
- Vogel S (1988). 'Les deux perfectifs des verbes "faire" et "devenir" dans les locutions en pashto.' *BSL* 1, 59–87.

Persian, Modern

P O Skjærvø, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Modern or New Persian (NPers. *fārsī*) is descended more or less directly from Middle Persian. Differences between Old, Middle, and New Persian in part reflect

the fact that these were official languages of the Achaemenid, Sasanian, and various modern-time dynasties of various local affiliations, respectively. The language is known from scattered remains from the 7th century C.E. on, while Persian literature emerged in the 9th century. Among the earliest manuscripts are a few texts from Chinese Turkestan.

New Persian is written in the Arabic script, with the exception of the Central Asian variant, Tajik, which uses the Cyrillic script. There is a large and old Judeo-Persian literature, written in the Hebrew alphabet, and a few short texts in Manichean and Syriac script.

There are several dialects of Modern Persian, among them the east-Iranian Khorasani dialect; Dari and Badakhshani spoken in Afghanistan; and Tajik (q.v.) spoken in Tajikistan and adjacent areas of Afghanistan and Xinjiang. Persian is in turn a member of the group of dialects spoken in western (Lorestan) and southwestern (Fars) Iran. Persian has also not been a homogeneous language throughout its history; rather, as the cultural centers moved about, the literary language was colored by the local varieties of Persian.

The study of Persian began in Persia itself probably already in the 13th to 14th centuries, but glossaries of obsolete and dialect words had been compiled as early as ca. 1000. Interest in ancient Iranian languages was kindled in Mughal India in the 16th century and resulted in several large dictionaries, which served as the basis for 19th- to 20th-century dictionaries.

In Europe, several Persian grammars had been written by 1700, partly based on Bible translations. About this time polyglot dictionaries, including Persian among their languages, also became common. The most famous early grammar is that of Sir William Jones (1st edn., 1771), founder of the Asiatick Society in Calcutta (1784).

Phonology

New Persian phonology continues that of Middle Persian with only few changes. Some early manuscripts mark intervocalic *b* and *d* as spirants β (<*f*> with triple dots) and δ , but the later standard language has only *b* and only *d* with a few exceptions (e.g., *godašt* → *gozašt*). The sounds γ and \dot{z} are common, but originate in non-Persian dialects (e.g., *rouyan* ‘oil,’ *vāže* ‘word’). In Arabic loan words, the typical Arabic sounds have been replaced with corresponding Persian ones (ح *ḥ* > *h*; ط *ṭ* > *t*; ث *ṭ* [θ], ص *ṣ* > *s*; ذ *ḏ* [δ], ض *ḏ*, and ظ *ẓ* > *z*). The Middle Persian vowel system: *a ā ē ī i u ū ō au ai* remained in early Modern Persian, except that *ē*, *ō* before nasals merged with *i*, *ū* early on in standard Persian. Later this merger took place in all positions, and eventually, this phonemic system based on vowel quantity distinctions was replaced by one based on vowel quality: *a* (≈ [æ]) *ā* (≈ [ɑ]) *e ī o ū ou ei* (with length as a secondary feature). The Classical Persian labialized velar fricative *xw* [x^o] is New Persian [x], but is still spelled <*xw*> (e.g., *xwad* > *x’od* ‘own’).

In modern colloquial standard Persian, further changes are taking place, among them *ān*, *ām* > *ūn*, *ūm*; loss of *h* and glottal stop before consonants, creating a new set of long vowels (*te:run* ‘Tehran’; *ba’d* [bæ’d] > [bæ:d] ‘afterward’ versus [bæd] ‘bad’); the assimilation of postvocalic *st* > *ss*; the reduction of the 3RD SING verb ending *-ad* to *-e* and the change of the 2ND PLUR ending *-īd* to *-īn*; and the reduction of the stem of several common verbs (*šaw* ‘become’ > *š-*, *gūy* ‘say’ > *g-*, etc., e.g., *mī-šav-ad* > *mī-š-e* ‘he/she/it becomes, it is possible,’ *mī-gūy-īd* > *mī-g-īn* ‘you [PL] say’).

Morphology and Syntax

Modern Persian has no grammatical gender (e.g., *ū* ‘he, she, it’) or cases, but has inherited the Middle Persian plural morphemes *-ān* (usually marked animate) and *-hā*. Arabic nouns often use the Arabic broken plural (*fe’l* ‘verb,’ plur. *af’āl*). Some Persian nouns have adopted the Arabic plural ending *-āt*, especially nouns in *-e* (<*-ag*), which have plural ending *-ejāt* (<*-aj-āt*; e.g. *mīve-jāt* ‘fruit’). Plural forms are not used after numerals, but classifiers are common, the unmarked *tā* ‘piece’ being the most common (*se jeld ketāb* or *se tā ketāb* ‘three volume/piece book’ = ‘three books’; *čand tā* or [colloquial] *čan dune* ‘how many?’ *se tā* or *se dune* ‘three’).

The indefinite marker is *-ī* ‘a (certain).’ The definite direct object is always marked, in Classical Persian by a variety of affixes, in modern Persian by *-rā* (colloquial often *-[r]o*, e.g., *to -rā ne-mī-šenās -am* ‘you-DO NEG-CONT-KNOW.PRES-1ST.SING’ = ‘I do not know you’; but *man* + *-rā* ‘I.DO’ > *marā* ‘me,’ coll. *man-o*), which can be combined with the indefinite article, *-ī-rā* ‘a certain.’ In colloquial Persian a referential definite marker is often used, e.g., *mard-é* ‘the man (we are talking about).’ The indirect object is marked in modern Persian by the preposition *be* ‘to,’ while in Classical Persian *-rā* was used, beside various other strategies. This was also the way of expressing possession (*ū-rā do bače and* ‘he.10 two child be.3RD.PL’ = ‘he has two children’; modern *do tā bače dār-ad* ‘two CLASS child have.PRES.-3RD.SING’ = ‘he he has two children’).

Adnominal constructions, possession, and adjectives, are expressed by the *ezafe* construction (*ketāb-e man* ‘book-CONN I’ = ‘my book,’ *ketāb-e bozorg* ‘book-CONN big’ = ‘a big book’). Possession can also be expressed by constructions such as *az ān-e man ast* ‘of that/those-CONN I.OBL be.3RD.SING’ = ‘it is mine’ or *māl-e man ast*, literally, ‘it is my possession.’ The *ezafe* is omitted after the indefinite article (*ketāb-i bozorg* book-INDEF big = ‘a big book’).

Relative clauses are introduced by the connector *ke*, which is preceded by *-ī* (attached to the noun) in restrictive clauses, e.g., (*mard-ī ke ketāb-am-rā bord* ‘man-REL.PART REL.CONJ book-I.OBL-DO take.away. PAST.3RD.SING’ = ‘the man who took my book’). The direct object particle may be added to the relative *-ī*, e.g., (*zan-ī-rā ke dī-rūz dīd-am* ‘woman-REL.PART-DO REL.CONJ yester-day see.PAST-1ST.SING’ = ‘the woman I saw yesterday.’ Anaphoric pronouns referencing the head noun are common, e.g., (*zan-ī dī-rūz dīd-am ke šouhar-eš dar jāng košt-e šod-e būd* ‘woman-INDEF yester-day see.PAST-1S.SING REL.CONJ husband-she.OBL in war kill.PAST-PERF become.PAST-PERF be.PAST.3RD-SING’ = ‘yesterday I saw a woman whose husband had been killed in the war.’ Note also constructions like *mard-ī dīd-am (ke) dāšt rāh mī-raft* ‘man-INDEF see.PAST-1ST.SING (CONJ) hold.PAST. 3RD.SING road PROG-go.PAST.3RD.SING’ = ‘I saw a man (that) he was walking the road’ = ‘I saw a man who was walking along’).

The verb system is based on three stems: present, past, and perfect (perfect participle = past stem + suffix *-e*; e.g., *kon-*, *kard*, *kard-e* ‘do’). The infinitive is made from the past stem (*kard-an* ‘to do’). To these stems are added personal endings and modal prefixes. The personal endings of the present and past tenses are the same, with the exception of the 3rd singular, which has no ending in the past tenses.

The most obvious feature distinguishing New Persian from its ancestors is the loss of the split-ergative (e.g., MPers. *ras-īd b-ēm* ‘arrive-PAST be.PRES.1ST.SING’ = ‘I arrived,’ *ras-īd* ‘he arrived > NPers. *ras-īd-am*, *ras-īd*; MPers. *man guft* ‘I.OBL say.PAST.3RD.SING’ = ‘I said’ > NPers. *gōft-am* ‘I did’; MPers. *ā-š guft* ‘then say.PAST.3RD.SING-he.ENCL.OBL’ = ‘then he said’ > NPers. *goft* ‘he said’ [in colloquial the 3rd singular enclitic pronoun may be added, *gōft-eš* ‘he said’]).

The perfect and pluperfect are formed with the extended past stem (perfect participle) in *-e* (*goft-ē-am* [colloquial *goftām*] ‘I have/had said,’ *goft-ē būd-am* ‘I had said’). Continuous tenses, including the perfect, take the prefix *mī-* (Class. Pers. also *hamē*). New progressive forms take the auxiliary *dār-* ‘hold’: *dār-am mī-rav-am* ‘hold.PRES-1ST.SING PROG-go.PRES.1ST.SING’ = ‘I am going, I am about to go.’

In Classical Persian the past tense takes the prefix *be-*, which in modern Persian is restricted to modal functions.

The future is formed with the verb *x’āstan* ‘wish’ and with a short form of the infinitive (*x’āh-ad boland šod* ‘future-3RD.SING tall become.SHORT-INF’ = ‘he is about to get up,’ different from *mī-x’āh-ad boland be-šav-ad* ‘he wishes to get up’ with the subjunctive). This construction can also be used

to mean ‘be about to’ (colloquial *mī-xās bolan š-e* [*mī-x’āst boland šav-ad*] ‘he was about to get up’).

The passive is formed with the auxiliary *šudan* ‘become’ (but MPers. ‘go’) and the perfect participle (*košt-e šod* ‘he was killed’). In earlier literature *āmadan* ‘come’ was often used instead of *šudan* (*nebešt-e mī-āmad* ‘it was written’).

Verbal system:

Present continuous: *mī-rav-am* ‘I go, I am going’

Present

subjunctive: *bē-rav-am* ‘(that) I go’

Past simple: *rāft-am* ‘I went’

Past continuous: *mī-raft-am* ‘I was going’

Perfect simple: *raft-ē-am* ‘I have gone’

Perfect continuous: *mī-raft-e-am* ‘I have (regularly) gone’

Pluperfect: *raft-ē būd-am* ‘I had gone’

Pluperfect continuous: *mī-raft-ē būd-am* ‘I would have gone’

Future: *x’āh-am raft* ‘I shall go’

Preverbs (local) are very common, and the meaning of the compound not always predicatable (e.g., *dāštan* ‘have, hold,’ *bar dāštan* ‘remove’). Verbal phrases are also very common, often comprising Arabic nouns and adjectives (e.g., *ettefāq* ‘incident’ + *oftādan* ‘fall’ = ‘happen’).

The local varieties of Persian have numerous variant forms, notably Afghan Dari and, especially, Tajik, which has been influenced by Turkic languages and has forms such as progressive present *karda-istoda-ast* ‘he is doing,’ inferential present *me-karda-ast* ‘he does (he says),’ presumptive *me-karda-gi-ast* ‘he appears to be doing, he probably does,’ etc.

Among syntactic features, we may note the following.

The *ezafe* construction can be used to connect extended qualifiers to head nouns, including prepositional phrases (*rāh-e be Qom* ‘the road to Qom’).

Passive constructions are agent-less (agents can only be expressed *ad hoc* in special phrases: *be-vasile-ye* ‘by means of,’ *az jāneb-e/taaraf-e* ‘from the side of,’ *be-dast-e* ‘at the hand of’).

The past continuous can express irrealis conditions (*agar mī-dān-est-am mī-goft-am* ‘if CONT-know-PAST-1ST.SING, CONT-tell.PAST-1ST.SING’ = ‘if I knew, I would tell’; *agar mī-dān-est-am goft-e būd-am* ‘if I knew, I would have told’; an alternative expression is *mī-dān-est-am ke mī-goft-am* ‘did I know, then I would say’).

The conjunction *ke* is used in a variety of functions and combinations (*vaqt-ī ke* ‘the time that’ = ‘when’; *be-jā-ye in ke man be-gūy-am* ‘instead of this that I should say’ = ‘instead of my saying’). There is no

indirect speech (*goft [ke] man mī-rav-am* ‘he said [that]: “I am going”’ = ‘he said he was going’; *porsīd ke šomā be-kojā mī-rav-īd* ‘he/she asked: “you [PL], to-where are you going?”’ = ‘he asked where they were going’; *mī-x’āst be-dān-ad ke be-mān-am yā na* [SUBJ-stay.PRES-1ST.SING or not] ‘he wished to know: “should I stay or not?”’ = ‘he wished to know whether I would stay/he should stay’).

Bibliography

Lazard G (1963). *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.

Lazard G (1989). ‘Le persan.’ In Schmitt R (ed.) *Compendium linguarum Iranicarum*. Wiesbaden: Reichert. 263–293.
 Lazard G (1992). *A grammar of contemporary Persian*. Lyon S A (trans.). Costa Mesa, CA: Mazda Publishers in association with Bibliotheca Persica.
 Lazard G (1993). *The origins of literary Persian*. Bethesda, MD: Foundation for Iranian Studies.
 Lazard G (1995). *La formation de la langue persane*. Paris: Peeters.
 Paul L (ed.) (2003). *Persian origins: Early Judaeo-Persian and the emergence of New Persian. Collected papers of the Symposium, Göttingen 1999*. Wiesbaden: Harrassowitz.
 Windfuhr G (1979). *Persian grammar: history and state of its study*. The Hague and New York: Mouton.

Persian, Old

E Tucker, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Old Persian was the native Iranian language of the Achaemenid Kings (522–330 B.C.), which they employed in their monumental inscriptions and foundation texts. Middle Persian and New Persian (Fārsī) are its direct descendants. Old Persian (OP) and Avestan together represent Old Iranian, that is, the earliest documented period of the Iranian language family, which is characterized by complex inflectional morphology inherited from the Indo-European parent language.

The remains of OP are not extensive, and most of the evidence belongs to the reigns of Darius I (522–486 B.C.) and his son Xerxes (486–465 B.C.). The later Achaemenids continued to compose short inscriptions in the same language, but there are indications that the spoken language was by this time evolving towards the Middle Persian stage. The meagre lexical data supplied by the OP texts is slightly enlarged by loanwords in the Persepolis Elamite Texts, Iranian words recorded by Greek authors, and Persian proper names in both literary and epigraphical sources from many areas of the ancient world.

Most Achaemenid inscriptions are trilingual, and the same text is repeated in Elamite, Akkadian, and OP. A simple form of cuneiform was invented to write OP, probably on the orders of Darius I, who wanted a Persian account of the events surrounding his own accession to accompany the relief and other texts at Mt. Bisitun in Media. There are 36 phonetic signs, including 3 for vowels; also 5 logograms, a set of numerals and a word divider. However, this specially devised writing system, which combines features of

an alphabet and a syllabary, renders the language only very imprecisely, and the interpretation of OP relies heavily on Avestan, Sanskrit and later Persian.

From an Indo-European perspective, OP shows the fundamental Indo-Iranian sound changes (IE **e, a, o, n, m* > *a*; **ē, ā, ō* > *ā*; IE labiovelars > velars, but palatals before an original front vowel; **s* > *š* after RUKI) and those changes that distinguish all Iranian languages from Indo-Aryan (**s* > *h* except before consonants; deaspiration of Indo-Iranian voiced aspirates; development of voiceless stops to spirants before consonants; dissimilation of dental clusters). Unlike Sanskrit, OP retains the Indo-Iranian diphthongs **ai*, **au* unchanged (OP *daiva-* ‘false god’, Av. *daēuua-*, Skt. *devā-*). IE **l*, **r* > OP *r*, and **l*, **r* probably > OP [ər], but the spelling is <a-r> initially, <-r> medially. Following a consonant **i*, **u* > OP *iy*, *uv* (OP *aniya-* ‘other’, Av. *āniia-*, Skt. *anyā-*).

Most notably, OP shows some SW Iranian dialect features. The outcome of the IE palatal stops **k̑*, **g̑*, **ǵh* is represented by *š*, *d*, *d*, probably all pronounced as spirants, in contrast to *s*, *z*, *z* in most Iranian languages (OP *viθ-* ‘(royal) household’, Av. *vīs-*, Skt. *viś-*; OP *drayah-* ‘sea’, Av. *zrayah-*, Skt. *jrāyas-*; OP *dasta-* ‘hand’, Av. *zasta-*, Skt. *hāsta-*). But IE **k̑u*, **ǵ(h)u* > OP *s*, *z* (OP *asa-* ‘horse’, Av. *aspa-*, Skt. *ásva-*), and **k̑n*, **ǵ(h)n* > *šn* (OP *baršnā* ‘in depth’). A series of changes result in new sibilants. IE **t̑* > **θy* > OP *šiy* (OP *hašiya-* ‘true’, Av. *haiθia-*, Skt. *satyā-*); IE **k̑w̑i*, **k̑i* > **cy* > OP *šiy* (OP *šiyav-* ‘to go’, Skt. *cyu-*); IE **tr* > **θr* > OP *ç* (possibly an affricate: OP *puça-* ‘son’, Av. *puθra-*, Skt. *putrá-*). Also IE **su*, **s̑u* > **hu*, **hv* > OP *u*, *uv* (*ubārta-* ‘well-borne’, Skt. *sūbhṛta-*). A number of words found in OP texts do not show the regular SW Iranian development (*vazrka-* ‘great’, *vīspa-* ‘all’, *xšāyaθiya-* ‘king’, etc.). They are traditionally

explained as loanwords from ‘Median’, but this is unverifiable.

Changes in OP final syllables have important consequences for inflectional morphology. Final **-t*, *-n*, *-h* are never written (*abara* for both **abarat*, **abaran* ‘he, they brought’, *pārsa* for **pārsah* ‘Persia’ nom. sg.); but syllables that ended in an original final vowel are treated differently, as both final **-a* and final **-ā* are written with an extra sign <-a>, probably indicating lengthening (*amariyatā* for **amariyata* ‘he died’; *pārsā* for **pārsā* instr. sg., and the same spelling for **pārsāt* abl. sg.).

OP nouns, adjectives, and pronouns inflect with three numbers (sg., dual and pl.) and there are three genders (masc., fem., neuter), but the eight inherited cases have been reduced to six. The forms of the Indo-Iranian dative have been lost and its functions taken over everywhere by genitive forms (*-ahyā*, *-ānām* in thematic stems, the most frequent type of nominal stem in OP). Ablative and instrumental have also merged; their inflections had become identical in the singular of nouns with vowel stems, but the demonstrative pronouns/pronominal adjectives possess a characteristic instrumental singular in *-nā* (*avanā*, ‘with/from that’). In feminine *ā*-stems, most of the singular cases have also become formally identical (gen.-dat., instr.-abl., locative, all in *-āyā*). The inflection of other stem-types is only partially attested, but some forms are remarkable (e.g., nom. sg. *pitā* ‘father’, gen.-dat. sg. *piça* < **pitrah*; from n-stems, acc. sg. *asmānam* ‘sky’).

The OP verb distinguishes three persons, three numbers; active and middle voices (but passive is expressed by a particular type of present stem in *-ya-* with active endings); and indicative, imperative, subjunctive and optative moods. Its tense system normally consists of a simple opposition between present and preterite forms based on the same stem, continuing inherited present vs. imperfect (*baratiy* ‘bears’, *abara* ‘bore’). Aorists and perfects are only preserved as relic forms, sometimes with a particular function (the sole perfect, *caxriyā*, is a perfect optative with irrealis value). The inherited augment *a-* is prefixed to all verb forms that indicate past time, including two that are formally optatives and indicate a habitual past action (*akunavaya(n)tā* ‘they would do’, *avājanīyā* ‘he used to kill’).

A new type of periphrastic past is built by means of the inherited past participle passive. It is intransitive (*paraitā* ‘they went off’) or passive (*haya idā krtā* ‘which was made here’). The agent is genitive-dative (*manā krtam* ‘done by me’ = ‘I did / have done’; *tayamaiy piça krtam āha* ‘what had been done by my father’). This construction is the ancestor of the Middle Persian past tense (*man kard* ‘I did’) and that of most New Iranian languages. The ppp. was also used in a construction with finite forms from root *kar-* ‘to do, to make’, which developed a potential value (*yātā krtam akunavam* literally ‘until I did it done’ = ‘until I succeeded in doing it’) and corresponds to the potentialis in Sogdian and Khotanese. The OP infinitive in *-tanaīy* is unparalleled elsewhere in IE, but is continued by the later Persian infinitive in *-tan/-dan*.

OP has a relative pronoun *haya-/taya-* that originated from a combination of the IE demonstrative **so*, *sā*, *tod* and the IE relative stem **yo-*, which were employed in correlative clauses in Indo-Iranian. In addition, this pronoun is used to connect qualifiers in a manner that prefigures the Persian *ezafe* (*Gaumāta haya maguš* ‘Gaumāta the Magian’). The inscriptions (particularly Bisitun) also abound in paratactic constructions of the type: *vašnā Auramazdāha Tigrām viyatarayāmā avadā avam karam tayam Nadi(n)tabairahyā ajanam vasiy Āciydiyahya māhyā XXVI raucabiš šakatā āha avašā hamaranam akumā* ‘By the will of Auramazdā we crossed the Tigris, there I defeated utterly that army of Nidintu-Bēl, of the month Āciyādiya 26 days were passed, then we made battle’.

Bibliography

- Kent R G (1953). *Old Persian: grammar, texts, lexicon* (2nd edn.). New Haven: American Oriental Society.
- Mayrhofer M (1978). *Supplement zur Sammlung der altpersischen Inschriften*. Vienna: Österreichische Akademie der Wissenschaften.
- Schmitt R (1989). ‘Altpersisch.’ In Schmitt R (ed.) *Compendium linguarum Iranicarum*. Wiesbaden: Reichert. 56–85.
- Schmitt R (1991). *The Bisitun inscriptions of Darius the Great: Old Persian text*. London: SOAS.
- Schmitt R (2000). *The Old Persian inscriptions of Naqsh-e Rostam and Persepolis*. London: SOAS.

Phoenician

J A Hackett, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Phoenician is a member of the Canaanite branch of the Northwest Semitic languages, closely related to Hebrew, Moabite, Ammonite, and Edomite. Phoenician was spoken both in the Levantine homeland and in the widespread Mediterranean colonies of the Phoenician commercial empire. Phoenicia itself is generally defined as the 60-mile long and 30-mile wide land area, from Acco to Tell Sukas south to north, and from the Mediterranean to the Lebanon Mountains, west to east (that is, the coast of modern Lebanon and part of the coast of modern Israel). It is scholarly convention to refer to this strip of land as Phoenicia after ca. 1200 B.C., the beginning of the Iron Age in the Levant. The ‘Sea Peoples’ (e.g., the Philistines) had forced the withdrawal of Egypt from ancient Canaan and had taken over the southern coastal region from them. The Sea Peoples do not seem to have carried their war to the northern coastal region, however, and so once the area was free of Egyptian control, the northern coastal cities became autonomous. They were never a single political entity, ‘Phoenicia,’ but rather a group of individual cities, although at any given time, one city was generally dominant over the others. The ancient Phoenician cities include Tyre, Sidon, Byblos, Beirut, Sarepta, and Arwad. The people of Phoenicia called themselves Canaanites or referred to themselves as the citizens of their particular city.

Again, by scholarly convention, we refer to the language of the inscriptions found in the cities along this coastal strip as ‘Phoenician’ from ca. 1200 B.C. onward, although the first inscriptions of any length unearthed so far date to ca. 1000. In fact, 10th-century Byblian inscriptions are written in a dialect slightly different from the Standard Phoenician of the rest of these inscriptions, but they are recognizably Phoenician all the same. 12th–11th century inscriptions that might also represent writing by Phoenicians are fragmentary or have not been found *in situ*, so that classification is difficult and dating must be paleographic: bronze arrowheads, for instance, probably from the Beqa^c Valley between the Lebanon and Antilebanon mountains, that are inscribed with personal names; inscribed clay cones from Byblos, also bearing personal names; the Nora fragment with parts of four words, written boustrophedon.

The alphabet and language of Phoenician inscriptions were the subject of scholarly debate already in the 18th century; by mid-century, both language and

alphabet were reasonably well deciphered. The texts are for the most part royal, funerary, or votive. They have been found in Syria as well, and all over the Mediterranean area: Asia Minor, Egypt, Greece, Spain, Cyprus, Sicily, Sardinia, Malta, Rhodes, the Balearic Islands. Punic, the dialect of the Phoenician colony at Carthage and of its own far-flung trading empire, is a development from Phoenician (Carthage – *Qart-hadasht or ‘New Town’ – was founded by Tyrians in the late 9th or early 8th century B.C.), and Punic inscriptions date from the 6th century B.C. until 146 B.C., when Carthage was destroyed by the Romans. After 146, Punic inscriptions are referred to as Neopunic, although it is the script that changes noticeably rather than the language. These late Punic inscriptions continue until the 4th–5th centuries A.D., when Latino–Punic inscriptions are attested. Punic inscriptions are known to us from all over North Africa, from the islands of the Mediterranean, and from France and Spain. The majority of the known Punic inscriptions are the hundreds of child-sacrifice votive inscriptions from North Africa.

Phoenician inscriptions are written in the Phoenician alphabet and until late Punic times are written entirely consonantly, so the vocalization of the language is reconstructed from comparative linguistics and from the few outside sources that include Phoenician words: for instance, Hebrew, Assyrian, Babylonian, and Greek writings (the Phoenician in these sources is mostly personal names), and the *Poenulus* of Plautus, which includes some passages in garbled Punic. In late Punic, there are sporadic uses of vowel letters (called *matres lectionis*, ‘mothers of reading’): *ʾaleph* to represent [ē] and [o], for instance, and *ayin* to represent [a].

Nominals in Phoenician are marked for gender and number (singular and plural, with rare duals) and occur in two ‘states’: the absolute (unbound) state and the bound state. The bound state is used for initial members of genitive chains called construct chains and for nouns before pronominal possessive suffixes. There is a definite article in Phoenician, initial *h-* plus doubling of the next consonant, as in Biblical Hebrew.

Several shifts in vowel pronunciation can be traced through the history of the language. The movement from [*ā] to [ō] between Proto-Northwest Semitic and Proto-Canaanite is known as the ‘Canaanite shift’; a later shift, occurring at least by the 8th century B.C., is the ‘Phoenician shift’: accented /a/ in originally open syllables becomes /o/. The diphthongs *-aw and *-ay collapse in Phoenician to [ō] and [ē], respectively. The [ō] < *-aw and the [ō] from the

Canaanite shift (*[ā]) merge, and by late Punic have become [ū]; this later shift is part of a proposed chain that sees *[u] pronounced [ü] or [i].

The verbal system of Phoenician follows the general Central Semitic pattern: a perfective *qatal* (> [qatol]) suffix conjugation; an imperfective *yaqtul* prefix conjugation; an imperative; active and passive participles; an infinitive (called the infinitive construct); and Phoenician uses (especially seen in the 8th-century Karatepe inscription from Asia Minor) the so-called infinitive absolute, actually an adverb, to represent any verb form needed in context, for instance imperative, or future or past tense.

Phoenician uses V-S-O word order in verbal clauses and makes much use of nominal or 'verbless' clauses. The verbal stems include the G stem (the *Grundstamm* or basic verb); the N stem, with prefixed *-n-*, which is passive/reflexive; the D 'intensive' stem (called D because the middle root consonant is doubled); the C or causative stem, called *Yiphil* because of the *y-* prefix; plus Gt and tD reflexive stems. There is also evidence of internal passives within the G, D, and C stems.

Bibliography

- Amadasi Guzzo M G (1996). 'Phoenician-Punic.' In Meyers E M (ed.) *Oxford encyclopedia of the archaeology of the Near East* 3. New York: Oxford. 317–324.
- Cross F M (1989). 'The invention and development of the alphabet.' In Senner W M (ed.) *The origins of writing*. Lincoln: University of Nebraska. 77–90.

- Fox J (1996). 'A sequence of vowel shifts in Phoenician and other languages.' *Journal of Near Eastern Studies* 55, 37–47.
- Friedrich J, Röllig W & Amadasi Guzzo M G (1999). *Analecta Orientalia 55: Phönizisch-punische Grammatik* (3rd edn.). Rome: Pontifical Biblical Institute.
- Garr W R (1985). *Dialect geography of Syria-Palestine, 1000–586, B. C. E.* Philadelphia: University of Pennsylvania.
- Ginsberg H L (1970). 'The Northwest Semitic languages.' In Mazar B (ed.) *The world history of the Jewish people* 2. Givatayim: Jewish History Publications. 102–124.
- Hackett J A (2004). 'Phoenician and Punic.' In Woodard R (ed.) *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press. 365–385.
- Harris Z S (1936). *American oriental series 8: A grammar of the Phoenician language*. New Haven, CT: American Oriental Society.
- Krahmalkov C R (2002). 'Phoenician.' In Kaltner J & McKenzie S L (eds.) *Beyond Babel: a handbook for Biblical Hebrew and related languages*. Atlanta: Society of Biblical Literature Press. 207–222.
- McCarter P K (1996). *Ancient inscriptions: voices from the Biblical world*. Washington, DC: Biblical Archaeology Society.
- Naveh J (1982). *Early history of the alphabet*. Jerusalem: Magnes.
- Segert S (1976). *A grammar of Phoenician and Punic*. Munich: C. H. Beck.
- Ward W A (1996). 'Phoenicia.' In Meyers E M (ed.) *Oxford encyclopedia of the archaeology of the Near East* 3. New York: Oxford. 313–317.

Pictish

W Nicolaisen, University of Aberdeen,
Aberdeen, UK

© 2006 Elsevier Ltd. All rights reserved.

Pictish was the language spoken by the Picts, inhabitants of the northeast of Scotland, roughly from the Forth-Clyde line to the Cromarty Firth, but possibly also further afield, including the Northern and Western Isles, from the early centuries A.D. until the middle of the 9th century when, as the result of the merger of the kingdoms of the Scotti and the Picti under Kenneth MacAlpine, it was replaced by Gaelic, which had reached Scotland from Ireland from approximately 500 A.D. onward. The Picts were known as *Picti* (or *Pecti*) to the Roman military, who interpreted their name in Latin terms as cognate with *pictus* 'painted'.

They were referred to by the neighboring Anglo-Saxons as *Pehtas*, *Pihtas*, *Pyhtas*, *Peohtas*, or *Piohtas*; by the Norsemen as *Péttar* or *Péttir* (as in *Pétlandsfjörðr* 'the Pentland Firth'); and in Middle Welsh as *Peith-wyr*, but it is not known what they called themselves. No sentence in their language has been recorded, and our main sources for Pictish are king lists, inscriptions, and, particularly, place names.

The nature and linguistic affiliation of Pictish has attracted attention for a long time, and this scholarly, and sometimes not so scholarly, pre-occupation with the language(s) of the Picts has led to a comparative neglect of other aspects of Scottish linguistic history and prehistory, especially when it comes to the analysis and interpretation of early place names. As far as Pictish is concerned, however, the fascination for its linguistic status has resulted in a large variety of

theories that have been offered, and often seriously defended, right to our own time. As recently as 1998, for example, Paul Dunbavin regarded the Picts as Finno-Ugric immigrants from the Baltic coast, basing his revolutionary conclusion, among other arguments, on the apparent derivation of certain Scottish river names, mentioned by Ptolemy about 150 A.D., from certain Finnish topographic terms. Apart from the absence of any documentary reference to Finno-Ugric people in Pictland, the proposed etymologies suffer from the frequently encountered flaw in such studies, the superficial equation of spellings reported almost 2000 years ago with modern forms of words in an otherwise unconnected language.

Whereas Dunbavin used toponymic, especially hydronymic, materials to support his proposal, Harald V. Sverdrup (1995), in the course of classifying and translating Pictish inscriptions, claimed “that it can be shown . . . that [Pictish] was neither a Celtic nor an Indo-European language but was distantly related to Caucasian languages,” dating the arrival of the initial settlers to the paleolithic transition before 7000 B.C. This would predate by several thousand years any other known nonmaterial evidence, including place names. It is the presumed enigmatic nature of Pictish that has led Sverdrup to the underlying readings, classification, and translation of the inscriptions.

A considerably earlier perception of Pictish as a non-Indo-European language comes from John Rhys, who in 1892, after discounting the Ugro-Finnish people (Lapps, Finns, and Estonians) and the Ligurians, felt “logically bound to inquire what Basque can do to help us to an understanding of the Pictish inscriptions.” However, 6 years later he revised his own theory by making it known that he no longer thought Pictish was related to Basque but rather to be pre-Indo-European (although not as old as the neolithic or mesolithic periods) that first came under p-Celtic influence from the Cumbrians south of the Forth-Clyde line. This change of direction did not stop J. B. Johnston, maintaining in 1934 his view (first expressed in 1892) that the river Urr in southwest Scotland derives from the Basque *ur* ‘water’, from falling into the same trap as Dunbavin.

One of the most outspoken opponents of a Celtic interpretation of the Pictish inscriptions was the Irish archeologist R. A. S. Macalister, who in 1922 expressed the view: “The most reasonable theory about the Picts was that they were survivals of the aboriginal pre-Celtic Bronze Age people. Certainly no attempt at explaining the Pictish inscriptions by means of any Celtic language could be called successful.” John Fraser, too, held in 1927 that, having

arrived before the Scots and the Britons, they must at one time have spoken a non-Indo-European language, although he took into account the later influence of (Scots) Gaelic and Brittonic.

Advocates for a non-Indo-European and often specifically of an anti-Celtic designation of Pictish represented a wide, fragmented variety of linguistic affiliations. In contrast, the pro-Celtic camp was divided into two opposing groups: those who regarded it as a q-Celtic language and those who regarded it as a p-Celtic language. Following such illustrious predecessors as George Buchanan (1582) and James Macpherson of Ossian fame (1763), Francis J. Diack (1944) was one of the strongest proponents of an uninterrupted Gaelic history of Scotland from the 1st century until today; the Gaelic nature of Pictish was asserted as recently as 1994 by Sheila McGregor.

The p-Celtic school also has a respectable pedigree in William Camden (1586) and Father Innes (1729). W. F. Skene (1868) declared Pictish to be neither Welsh nor Gaelic but “a Gaelic dialect partaking largely of Welsh forms.” One of the first scholars to put the p-Celtic nature of Pictish on a sound footing was Alexander Macbain (1891–1892) in his survey of ‘Ptolemy’s geography of Scotland’; his stance was strongly supported by W. J. Watson (1904, 1921, 1926), mainly on the basis of place-name evidence. This is also at the heart of Kenneth H. Jackson’s (1955) overview of ‘The Pictish language,’ in the course of which he presents the first maps of linguistic Pictland based on the distribution of such place-name elements as *pett* (Pit-), *aber*, *carden*, *lanerc*, *pert*, and *pevr*. He also suggested, however, that there may have been two Pictish languages, one the language of the pre-Indo-European inhabitants, the other the Gallo-Brittonic tongue of Iron Age invaders. W. F. H. Nicolaisen acknowledged the presence of non- or pre-Indo-European place names in Pictish territory, but did not regard them as Pictish. In her repudiation of Jackson’s two Pictishes, in her thorough investigation of *Language in Pictland*, Katherine Forsyth (1997), a specialist in Ogham inscriptions, mustered some very persuasive arguments against Jackson’s construct and, although it is always risky to call anything ‘definitive,’ her conclusion that the Picts were “as fully Celtic as their Irish and British neighbors” is difficult to dispute, and it is good to see Pictish placed where it belongs beside other p-Celtic languages such as Cumbric, Welsh, Cornish, Breton, and Gaulish; by implication, the firm ascription of Pictish in this linguistic grouping adjudges the Scotti to have brought the Gaelic language with them from Ireland, a consequence of fundamental importance in a long-running debate.

Bibliography

- Cox R A V (1997). 'Modern Gaelic reflexes of two Pictish words.' *Nomina* 20, 47–58.
- Cummins W A (1998). *The age of the Picts*. Stroud, UK: Sutton.
- Diack F C (1920–1921/1922). 'Place-names of Pictland.' *Revue Celtique* 38, 109–132; 39, 125–174.
- Diack F C (1944). *The inscriptions of Pictland*. Aberdeen, UK: Third Spalding Club.
- Dunbavin R (1998). *Picts and ancient Britons: an exploration of Pictish origins*. Long Eaton, UK: Third Millennium Publication.
- Forsyth K C (1996). *The Ogham inscriptions of Scotland: an edited corpus*. Ph.D. diss., Harvard University.
- Forsyth K (1997). *Language in Pictland: the case against non-Indo-European Pictish*. Utrecht: de Keltische Draak.
- Fraser J (1927). 'The question of the Picts.' *Scottish Gaelic Studies* 2, 172–201.
- Fraser J (1942). 'Pet(t) in place-names.' *Scottish Gaelic Studies* 5, 67–71.
- Henderson I (1967). *The Picts*. London: Thames and Hudson.
- Henry O (ed.) (1992). *The worm, the germ and the thorn*. Balgavies, UK: The Pinkfoot Press.
- Jackson K H (1955). 'The Pictish language.' In Wainwright F T (ed.) *The problem of the Picts*. Edinburgh, UK: Nelson. 129–166.
- Johnston J B (1934). *Place-names of Scotland* (3rd edn.). London: John Murray.
- Macalister R A S (1940). 'The inscriptions and language of the Picts.' In Ryan J (ed.) *Féil-scribbian Eóin Mhic Néill: essays and studies presented to Professor Eóin MacNeill*. Dublin, Ireland: At the Sign of the Three Candles. 184–226.
- Macbain A (1891–1892). 'Ptolemy's geography of Scotland.' *Transactions of the Gaelic Society of Inverness* 128, 267–288.
- MacGregor S (1994). 'The abers of Perthshire.' *Pictish Arts Society Journal (spring)*, 12–19.
- Nicolaisen W F H (1996). *The Picts and their place-names*. Rosemarkie, UK: Groam House.
- Nicolaisen W F H (2001). *Scottish place-names: their study and significance* (rev. edn.). Edinburgh, UK: John Donald.
- Nicolaisen W F H (2004). 'A gallimaufry of languages.' In Van Nahl A, Elmevik L & Brink S (eds.) *Namenwelten: orts and personennamen in historischer sicht*. Berlin and New York: de Gruyter. 233–240.
- Nicoll E H (ed.) (1995). *A Pictish panorama*. Balgavies, UK: The Pinkfoot Press.
- Rhys J (1891–1892). 'The inscriptions and language of the northern Picts.' *Proceedings of the Society of Antiquaries of Scotland* 126, 263–351.
- Rhys J (1897–1898). 'A revised account of the inscriptions of the northern Picts.' *Proceedings of the Society of Antiquaries of Scotland* 132, 324–398.
- Ritchie A (1989). *Picts*. Edinburgh, UK: H. M. Stationery Office.
- Skene W F (1868). *Four ancient books of Wales* (Vol. 1). Edinburgh, UK: Edmonston and Douglas.
- Sutherland E (1994). *In search of the Picts*. London: Constable.
- Sverdrup H (1995). *Reports in ecology and environmental engineering 4: Classifying and translating inscriptions in the Pictish language*. Lund.
- Taylor S (1994). 'Generic-element variation with special reference to eastern Scotland.' *Nomina* 20, 5–22.
- Wainwright F T (ed.) (1955). *The problem of the Picts*. Edinburgh, UK: Nelson.
- Watson W J (1904). *Place-names of Ross and Cromarty*. Inverness, UK: The Northern Counties Printing and Publishing Company.
- Watson W J (1921). 'The Picts: their original position in Scotland.' *Transactions of the Gaelic Society of Inverness* 80, 240–261.
- Watson W J (1926). *The history of the Celtic place-names of Scotland*. Edinburgh, UK: Blackwood.

Pidgins and Creoles

S Romaine, Oxford University, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Definitions

European colonization during the 17th to 19th centuries created a classic scenario for the emergence of new language varieties called pidgins and creoles out of trade between the native inhabitants and

Europeans. The term 'pidgin' is probably a distortion of English *business* and the term 'creole' was used in reference to a nonindigenous person born in the American colonies, and later used to refer to customs, flora, and fauna of these colonies. Many pidgins and creoles grew up around trade routes in the Atlantic or Pacific, and subsequently in settlement colonies on plantations, where a multilingual work force comprised of slaves or indentured immigrant laborers needed a common language. Although European

colonial encounters have produced the most well known and studied languages, there are examples of indigenous pidgins and creoles predating European contact such as Mobilian Jargon (Mobilian), a now extinct pidgin based on Muskogean (Muskogee), and widely used along the lower Mississippi River valley for communication among native Americans speaking Choctaw, Chickasaw, and other languages (see **Mobilian Jargon**).

The study of pidgins and creoles raises fundamental questions about the evolution of complex systems, since pidgins, in particular, have been traditionally regarded as simple systems *par excellence*. The usual European explanation given for the simplicity, and lack of highly developed inflectional morphology in particular, was that it reflected primitiveness, native mental inferiority, and the cognitive inability of the natives to acquire more complex European languages. Thus, for example, Churchill (1911: 23) on Bislama, the pidgin English spoken in Vanuatu: “the savage of our study, like many other primitive thinker, has no conception of being in the absolute; his speech has no true verb ‘to be’” (see **Bislama**).

Hampered by negative attitudes for many years, scholars ignored pidgins and creoles in the belief that they were not ‘real’ languages, but were instead bastardized, corrupted, or inferior versions of the European languages to which they appeared most closely related. Although scholars still do not agree on how to define pidgins and creoles, or the nature of their relationship to one another, most linguists recognize such a group of languages, whether defined in terms of shared structural properties and/or socio-historical circumstances of their genesis. Striking similarities across pidgin and creole tense-mood-aspect (TMA) systems were noted by some of the earliest scholars in the field such as Hugo Schuchardt, generally regarded as the founding father of creole studies. TMA marking became a focal point of debate among creolists as a result of the bioprogram hypothesis (Bickerton, 1981, 1984), according to which creoles held the key to understanding how human languages originally evolved many centuries ago. This theory led not only to an increase in research on these languages, but also a great deal of attention from scholars in other fields of linguistics, such as language acquisition and related disciplines such as cognitive science.

Classifying Pidgins and Creoles

The standard view that pidgins and creoles are mixed languages with the vocabulary of the superstrate (also called the lexifier or base language) and the grammar of the substrate (the native languages of the groups in

contact) has been the traditional basis for classifying these languages according to their lexical affiliation. English-lexicon pidgins and creoles such as Solomon Islands Pijin spoken in the Solomon Islands or Jamaican Creole English (Southwestern Caribbean Creole English) in Jamaica comprise a group of languages with lexicons predominantly derived from English. Haitian Creole French and Tayo, a French creole of New Caledonia, are French-lexicon creoles drawing most of their vocabulary from French. Such groupings are, however, distinctly different from the genetically-based language families established by the comparative historical method. Pidgins or creoles as a group are not genetically related among themselves, although those with the same lexifier usually are.

There is a great deal of variation in terms of the extent to which a particular pidgin or creole draws on its lexifier for vocabulary, and a variety of problems in determining the sources of words, due to phonological restructuring. Compare the lexical composition of Sranan and Saramaccan, two of six English-lexicon creoles spoken in Surinam, in what was formerly the Dutch-controlled part of Guyana. About 50% of the words in Saramaccan are from English (e.g., *wáka* ‘walk’), 10% from Dutch (e.g., *strei* ‘fight’ < strijd), 35% from Portuguese (e.g., *disá* ‘quit’ < deixar), and 5% from the African substrate languages (e.g., *totómbotí* ‘woodpecker’. By contrast, only 18% of Sranan words are English in origin, with 4.3% of African origin, 3.2% of Portuguese, 21.5% of Dutch; 4.3% could be derived from either English or Dutch. Innovations comprise another 36%, and 12.7% have other origins. African words are concentrated in the semantic domains of religion, traditional food, music, diseases, flora, and fauna. Words from the other languages do not concentrate in particular semantic domains. Numbers, for instance, draw on both English and Dutch. Sranan and Saramaccan are not mutually intelligible, and neither is mutually intelligible with any of the input languages. Other languages show a more equal distribution between two main languages, such as Russenorsk, a pidgin once spoken along the Arctic coast of northern Norway from the 18th until the early 20th century. Its vocabulary is 47% Norwegian, 39% Russian, 14% other languages including Dutch (or possibly German), English, Saami, French, Finnish, and Swedish (see **Russenorsk**).

Many creoles, like Lesser Antillean (Lesser Antillean Creole French), a French-based creole spoken in the French Antilles, started out with a far more mixed lexicon than they possess today. Where contact with the main European lexifier was permanently terminated, as in Surinam, the lexicon retains a high degree of mixture to the present day; where such contact

continued, as in the Lesser Antilles, items from the main lexifier tended gradually to replace items from other sources. Depending on the circumstances, a creole may adopt more items from the superstrate language due to intense contact. In Tok Pisin spoken in Papua New Guinea, some of the 200 German elements as well as words from indigenous languages, are now being replaced by English words. Thus, *beten* (German 'pray') is giving way to English *pre*, and Tolai (Kuanua) *kiau* to English 'egg' (see **Tok Pisin**).

Relationships between Pidgins and Creoles

The question of the genetic and typological relationship between pidgins and creoles and the languages spoken by their creators continues to generate controversy. Pidgins and creoles challenge conventional models of language change and genetic relationships because they appear to be descendants of neither the European languages from which they took most of their vocabulary, nor of the languages spoken by their creators. The conventional view of the languages and their relationship to one another found in a variety of introductory texts (Hall, 1966; Romaine, 1988) has been to assume that a pidgin is a contact variety restricted in form and function, and native to no one, which is formed by members of at least two (and usually more) groups of different linguistic backgrounds, e.g., Krio in Sierra Leone (see **Krio**). A creole is a nativized pidgin, expanded in form and function to meet the communicative needs of a community of native speakers, e.g., Haitian Creole French.

This perspective regards pidginization and creolization as mirror image processes and assumes a prior pidgin history for creoles. This view implies a two-stage development. The first involves rapid and drastic restructuring to produce a reduced and simplified language variety. The second consists of elaboration of this variety as its functions expand, and it becomes nativized or serves as the primary language of most of its speakers. The reduction in form characteristic of a pidgin follows from its restricted communicative functions. Pidgin speakers, who have another language, can get by with a minimum of grammatical apparatus, but the linguistic resources of a creole must be adequate to fulfill the communicative needs of human language users.

The degree of structural stability varies, depending on the extent of internal development and functional expansion the pidgin has undergone at any particular point in its life cycle. Creolization can occur at any stage in the development continuum from rudimentary jargon to expanded pidgin. If creolization occurs at the jargon stage, the amount of expansion will

be more considerable than that required to make an expanded pidgin structurally adequate. In some cases, however, pidgins may expand without nativization. Where this happens, pidgins and creoles may overlap in terms of the structural complexity, and there will be few, if any, structural differences between an expanded pidgin and a creole that develops from it. Varieties of Melanesian Pidgin English (a cover term for three English-lexicon pidgins/creoles in the southwest Pacific comprising Tok Pisin, Solomon Islands Pijin and Vanuatu Bislama) are far richer lexically and more complex grammatically than many early creoles elsewhere. Their linguistic elaboration was carried out primarily by adult second language speakers who used them as lingua francas in urban areas. Creolization is thus not a unique trigger for complexity, and the 'same' language may exist as both pidgin and creole.

Debate continues about the role of children vs. adults in nativization and creolization. Other scholars have emphasized the discontinuity between creoles and pidgins on the basis of features present in certain creoles not found in their antecedent pidgins. They argue that ordinary evolutionary processes leading to gradual divergence over time may not be applicable to creoles. Instead, creoles are 'born again' nongenetic languages that emerge abruptly *ab novo* via a break in transmission and radical restructuring (Thomason and Kaufman, 1988).

Origins

Because pidgins and creoles are the outcome of diverse processes and influences in situations of language contact where speakers of different languages have to work out a common means of communication, competing theories have emphasized the importance of different sources of influence. Few creolists believe that one theory can explain everything satisfactorily, and there are at least four theories accounting for the genesis of creoles: substrate, superstrate, diffusion, and universals.

Substrate

The substrate hypothesis emphasizes the influence of the speakers' ancestral languages. Structural affinities have been established between the languages of West Africa and many of the Atlantic creoles. Scholars have also documented substantial congruence between Austronesian substratum languages (see **Austronesian Languages**) and Pacific pidgins as compelling evidence of the historically primary role of Pacific Islanders in shaping a developing pidgin in the Pacific. Substrate influence can be seen in the pronominal systems of Melanesian Pidgin English

such as the personal pronouns in Tok Pisin. The forms are rather transparently modeled after English, yet incorporate grammatical distinctions not found in English, but widely present in the indigenous languages forming the substrate.

| Personal pronouns in Tok Pisin | | | |
|--------------------------------|-----------|-------------|--|
| | singular | | plural |
| first person | <i>mi</i> | 'I' | <i>mipela</i> 'we' (exclusive) <i>yumi</i> 'we' (inclusive) |
| second person | <i>yu</i> | 'you' | <i>yupela</i> 'you' |
| third person | <i>em</i> | 'he/she/it' | <i>ol</i> 'they' |

Almost all Oceanic languages distinguish between inclusive (referring to the speaker and addressee(s), 'I + you') and exclusive first-person pronouns (referring to the speaker and some other person(s), 'I + he/she/it/they'). Thus, *yumi* consists of the features [+speaker, +hearer, +other] and *mipela*, [+speaker, –hearer, +other]. There are also dual and trial forms, e.g., *yumitupela* 'we two (inclusive)', i.e. [+speaker, +hearer, –other], *mitripela* 'we three (exclusive)', etc., although these distinctions are not always made consistently. As English provides no lexical forms for the inclusive/exclusive and dual distinctions or *you* plural, these are created by forming a compound from *you* + *me* to give *yumi* and *yumitupela*, and by using the suffix *-pela* ('fellow') to mark plurality in *yupela*. The third-person singular form *em* is derived from the unstressed third person singular *him* and the third person plural form *ol* from *all*.

A more controversial variant of the substrate hypothesis is incorporated into the notion of relexification, a process that applies to the words/structures of substrate language and matches them with phonological representations from the lexifier language. Haitian Creole French *gade* shares some meanings with the French verb *garder* 'to watch over/take care of/to keep', from which it derives its phonetic form, but it has an additional meaning 'to take care of/defend oneself'. The semantics of *gade* is very similar to that of the substrate Fongbe (Fon-Gbe) verb *kpón* 'to watch over/take care of/to keep/to look'. Haitian Creole French *gade* also means 'to look', while in French that meaning is expressed by *regarder*. These similarities have led some scholars to regard Haitian Creole French as a French relexification of African languages of the Ewe-Fon (or Fongbe) group (Lefebvre, 1998).

Superstrate

The superstrate hypothesis traces the primary source of structural features to nonstandard varieties of the lexifiers, and to evolutionary tendencies already observable in them (Chaudenson, 1992). According to this scenario, early plantation slaves acquired a

normally transmitted variety of the lexifier directly from Europeans, but this imperfectly acquired variety was subsequently diluted over time as successive generations of slaves learned from other slaves rather than from Europeans. Creoles thus represent gradual continuous developments with no abrupt break in transmission from their lexifiers. This evidence eliminates the assumption of a prior pidgin history and accepts creoles as varieties of their lexifiers rather than as special or unique new languages. That is, there are no particular linguistic evolutionary processes likely to yield (prototypical) creoles; they are produced by the same restructuring processes that bring about change in any language. Creoles are neither typologically nor genetically unique, but 'advanced varieties' of the lexifiers.

Linguistic evidence supporting this hypothesis can be found in morphemes or constructions chosen for specific grammatical functions that start from models available in the lexifiers. Haitian Creole French *m pu alle* 'I will go' may not be a totally new and radical departure from French but could instead be derived from regional French *je suis pour aller*.

Diffusion

Another explanation for some of the similarities among pidgins and creoles is diffusion of a pre-existing pidgin. According to this hypothesis, a pre-existing English or French pidgin was transplanted from Africa rather than created anew independently in each territory. Support for this hypothesis can be found in historical evidence that sailors diffused not only words with nautical origins from one part of the world to another, but also items that were more generally part of regional and nonstandard usage. Thus, *capsize* was probably originally a nautical term meaning 'to overturn a boat'. Today, *kapsaitim* in Melanesian Pidgin English means 'to spill or overturn anything'. Traders, missionaries, and early settlers were also responsible for diffusing certain elements. Words from Portuguese such as *savvy* (<*sabir* 'to know/understand', first attested in 1686) are found widely around the world. Scholars have traced the paths of diffusion of so-called worldwide features found in Anglophone pidgins and creoles from the Atlantic to Pacific (Baker and Huber, 2001). Words from indigenous languages are also widespread, e.g., African *nyam* 'eat/food' and Hawaiian *kanaka* 'person/man', a term that came to be used, often derogatorily, to refer to Pacific Islanders.

Universals

This theory actually comprises a variety of sometimes opposing viewpoints because universals have been

conceived of in a variety of ways within different theoretical perspectives. Its central assumption is that creoles are more similar to one another than the languages to which they are otherwise most closely related due to the operation of universals. Although it has become fashionable to refer to a common creole syntax or creole prototype, not all creolists agree on the nature or extent of the similarities or the reasons for them. If creoles form a synchronically definable class, then there should be more similarities between Haitian Creole French and Guyanese Creole English than between Haitian Creole French and French, or between Guyanese Creole English and English. One kind of universalist claim is that creoles reflect more closely universal grammar and the innate component of the human language capacity. Another, however, is grounded within a different notion of universals derived from crosslinguistic typology and theories of markedness. The observation that creoles tend to be isolating languages even when the contributing languages show a different typology has a long history predating modern typological theories. Kituba, for example, emerged almost exclusively from contact among Bantu languages that are agglutinative.

The notion of creoles as the simplest instantiation of universal grammar is at the heart of Bickerton's (1981) bioprogram hypothesis, which applies to radical creoles, i.e., those that have undergone a sudden creolization without further major superstrate influence. It is based to a large extent on similarities between Hawai'i Creole English, Guyanese Creole English, Haitian Creole French, and Sranan. Evidence from Hawai'i Creole English has been the cornerstone of the bioprogram because creolization has been more recent there than in many other cases, and because the language lacked an African substrate, yet was strikingly similar to other creoles (see **Hawaiian Creole English**). This similarity is explained by assuming that creoles represent a retrograde evolutionary movement to a maximally unmarked state.

Bickerton (1981) proposed a list of 13 features shared by creoles that were not inherited from the antecedent pidgins, and therefore must have been created by children as a result of the bioprogram.

1. Focused constituents are moved to sentence initial position, e.g., Haitian Creole French *se mache Jan mache al lekòl* 'John walked to school'.
2. Creoles use a definite article for presupposed specific noun phrases, indefinite articles for asserted specific noun phrases, and zero for nonspecific noun phrases. Hawai'i Creole English uses definite article *da* for presupposed specific noun phrases, e.g., *she wen go with da teacher* 'she

went with the teacher', indefinite article *one* typically for first mention, e.g., *he get one white truck* 'he has a white truck', and no article or marker of plurality for other noun phrases, e.g., *young guys they no get job* 'Young people don't have jobs'.

3. Three preverbal morphemes express tense (anterior), mood (irrealis), and aspect (durative) in that order, e.g., Haitian Creole French *li te mache* 'he walked', *l'av(a) mache* 'he will walk', *l'ap mache* 'he is walking'.
4. Realized complements are either unmarked or marked with a different form than the one used for unrealized complements, e.g., Mauritian Creole French (Morisyen) *il desid al met posòb ladh* 'she decided to put a fish in it' vs. *li ti pe ale aswar pu al bril lakaz sa garsòb-la me lor sime ban dayin lin atake li* 'He would have gone that evening to burn the boy's house, but on the way he was attacked by witches'.
5. Creoles mark relative clauses when the head noun is the subject of the relative clause, e.g., Hawai'i Creole English *some they drink make trouble* 'Some who drink make trouble'.
6. Nondefinite subjects, nondefinite verb phrase constituents, and the verb must all be negated in negative sentences, e.g., Guyanese Creole English *non dag na bait non kyt* 'no dog bit any cat'.
7. Creoles use the same lexical item for both existentials and possessives, e.g., Hawai'i Creole English *get one wahine she get one daughter* 'There is a woman who has a daughter'.
8. Creoles have separate forms for each of the semantically distinct functions of the copula (i.e., locative and equative), e.g., Sranan *a ben de na ini a kamra* '(s)he was in the room.' vs. *mi na botoman* 'I am a boatman'.
9. Adjectives function as verbs, e.g., Jamaican Creole English *di pikni sik* 'the child is sick'. This function explains the absence of the copula in this construction.
10. There are no differences in word order between declaratives and questions, e.g., Guyanese Creole English *i bai di eg dem* means 'he bought the eggs' or 'did he buy the eggs?', depending on intonation.
11. Questions particles are optional and sentence final, e.g., Tok Pisin *yu tok wanem?* 'what did you say'. Question words are often bimorphemic, e.g., Haitian Creole French *ki kote* 'where' (French *qui côté* 'which side'), and Tok Pisin *wanem* 'which/what' (English *what name*).
12. Formally distinct passives are typically absent, e.g., Jamaican Creole English *dem plaan di tree* 'they planted the tree' vs. *di tree plaan* 'the tree was planted'.

13. Creoles have serial verb constructions in which chains of two or more verbs have the same subject, e.g., Nigerian Pidgin English (Pidgin, Nigerian) *dem come take night carry di wife, go give di man* 'They came in the night and carried the woman to her husband'.

There are also many similarities in the source morphemes used by creoles to express these distinctions. The semantics of the grammatical morphemes are highly constant as are their etymologies; in almost all cases, they are drawn from the superstrate language. The indefinite article is usually derived from the numeral 'one', the irrealis mood marker from a verb meaning 'go', the completive marker from a verb meaning 'finish', the irrealis complementizer from a reflex of 'for', etc.

Support for the uniqueness of these features to creoles is, however, weakened by the existence of some of the same traits in pidgins as well as in the relevant substrates and superstrates. The relexification hypothesis argues that the typological traits of Haitian Creole French display more in common with those of the substrate language Fongbe than with French. If so, then the supposed creole typology results from the reproduction of substratum properties rather than from the operation of universals. Bimorphemic question words are also found in many of the African substrate languages, and English has *what time* 'when', *how come* 'why', etc. It is also well within the norms of colloquial French and English to use intonation rather than word order to distinguish questions from declaratives, e.g., *you're doing what?* The absence of passives may also reflect the lack of models in some of the substrate and superstrate languages.

Closer study of the particulars of individual TMA systems in creole languages has engendered increasing dissatisfaction with the bioprogram hypothesis (Singler, 1990). For one thing, the claims were originally formulated on the basis of data from creoles whose superstrate languages are Indo-European. Secondly, it is also unclear how much creole TMA systems might have changed over time after creolization. The bioprogram assumes that the creoles in question have not departed from their original TMA prototype and that the present day systems provide evidence of relevance for its operation. Thirdly, even the defining languages do not conform entirely to predictions on closer examination. The TMA system of Hawai'i Creole English is not crosslinguistically unique or even unusual; the overwhelming majority of its TMA categories are common in languages of world (Velupillai, 2003). More detailed investigations of historical evidence indicate that Bickerton's

scenario of nativization bears little resemblance to what actually happened in Hawai'i (Roberts, 2000).

The typology of creoles might also be largely a result of parameter settings typical of languages with low inflectional morphology. Thus, features such as preverbal TMA markers, serial verbs, and SVO word order fall out more generally from lack of inflections and unmarked parametric settings. McWhorter (1998) attempts to vindicate creoles as a unique typological class by proposing a diagnostic test for 'creolity' based not on specific shared structural features such as TMA markers, serial verbs, etc., but on a combination of three traits resulting from a break in transmission: little or no use of inflectional affixation, little or no use of lexical tone, and semantically regular derivational affixation. McWhorter's explanation for why these traits cluster essentially reiterates the conventional assumption that pidgins are languages that have been stripped of all but the bare communicative necessities in order to speed acquisition. Because creoles are new languages that emerge from pidgins, they have not had the time to develop many of the complexities found in other languages that have developed gradually over a much longer time period. Thus, he predicts that features such as ergativity, a distinction between alienable and inalienable possession, switch reference marking, noun class or grammatical gender marking, etc. will never be found in creoles. This theory means that not only are creoles typologically unique, but also that they are the simplest languages. Those who stress the role of substrate influence and relexification, however, have argued that the reason why these features do not surface in creoles even where they are present in the substrate is because there are no appropriate phonetic strings in the superstrate to match them with.

The question of how to measure simplicity and complexity is theory-dependent and therefore controversial. McWhorter's (2001) complexity metric is based on degree of overt signalling of various phonetic, morphological, syntactic, and semantic distinctions. From this perspective, a phonemic inventory can be considered more complex if it contains more marked members than some other. Markedness is interpreted in terms of frequency of representation among the world's languages. Ejectives and clicks are more marked than ordinary consonants because they occur less frequently. The presence of rarer sounds in an inventory also presupposes the existence of more common or less marked ones. However, there may be other dimensions of simplicity/complexity to consider, such as syllable/word structure. Much less is known about the phonology of pidgins and creoles than about their syntax and lexicon. Syntax

is rendered more complex by the additional of rules that make it more difficult to process, e.g., different word orders for main and subordinate clauses. Inflectional marking is assumed to be more difficult than the use of free morphemes. However, there is no universally accepted account of syntactic rules nor an agreed theory of processing. Semantically, creoles are more transparent and adhere more closely to the principle of one form–one meaning.

There are problems with this view too, because creoles do not share their features universally or exclusively. There are examples of noncreole languages with the assumed typical creole-like features, and some examples of languages with no known creole history that are less complex than some creoles. Given that language change may also lead to simplification, some languages that are older than creoles may also be less complex than creoles. Similarities among creoles may be the result of chance similarities among unrelated substrates. Although the absence of inflection is perhaps the most often cited typological feature of creoles, it may be the accidental result of limited typological spread of the contributing languages.

Yet another interpretation of the universalist approach involves the assumption that common processes of restructuring apply in situations of language contact to produce common structural outcomes. The effects of contact may operate to differing degrees depending on the social context, e.g., number and nature of languages involved, extent of multilingualism, etc. The fact that pidgins and creoles share some structural features with each other and with other language varieties that are reduced in function such as koines, learner varieties, etc., indicates that the same solutions tend to recur to some degree wherever acquisition and change occurs, regardless of contact, but especially in cases of contact. The entities called pidgins and creoles are salient instances of the processes of pidginization and creolization respectively, although they are not in any sense to be regarded as unique or completed outcomes of them. From this point of view, pidgins represent a special or limiting case of reduction in form resulting from restriction in use.

This statement brings us back to the position that the only thing special about creoles is the sociohistorical situation of language contact in which they emerge. Even that may not be so special when we consider the history of so-called normal languages, most of which are hybrid varieties that have undergone restructuring to various degrees depending on the circumstances. Even ‘normal’ languages such as English have been shaped by heavy contact with non-Germanic languages and thus can be thought of as

having more than one parent. If universal grammar is a mental construct, or an innate predisposition to develop grammar, then in so far as there is no psychological continuity between the mental representations of one generation of speakers of a language and the next, all grammars are created anew each generation. There will always be a certain amount of discontinuity between the grammars of parents and children, and acquisition is always imperfect. Thus, the supposed dichotomy between normal and abrupt transmission is spurious because normal transmission is in fact abrupt.

Directions for Future Research

Resolution of some of the debates about pidgins and creoles, their origins, and their relationships to one another as well as to the languages spoken by their creators is hampered by lack of knowledge of the relevant substrate languages as well as insufficient knowledge of the history of the nonstandard varieties of European languages that formed the lexifiers. There are few detailed grammatical descriptions of pidgins and creoles available for sophisticated typological analysis. More sociohistorical research is also needed. Earlier scholarship often overstated the similarities among creoles and ignored key properties unique to individual ones.

Bibliography

- Baker P & Huber M (2001). ‘Atlantic, Pacific, and world-wide features in English-lexicon contact languages.’ *English World Wide* 22(2), 157–208.
- Bickerton D (1981). *Roots of language*. Ann Arbor: Karoma.
- Bickerton D (1984). ‘The language bioprogram hypothesis.’ *Behavioral and Brain Sciences* 7, 173–221.
- Bickerton D (1988). ‘Creoles languages and the bioprogram.’ In Newmeyer F J (ed.) *Linguistics: the Cambridge survey 2: Linguistic theory: extensions and implications*. Cambridge: Cambridge University Press. 268–284.
- Bickerton D & Muysken P (1988). ‘A dialog concerning the linguistic status of creole languages.’ In Newmeyer F J (ed.) *Linguistics: the Cambridge survey 2: Linguistic theory: extensions and implications*. Cambridge: Cambridge University Press. 302–306.
- Chaudenson R (1992). *Des îles, des hommes, des langues: essais sur la créolisation linguistique et culturelle*. Paris: L’Harmattan.
- Churchill W (1911). *Beach-La-Mar, the jargon trade speech of the Western Pacific*. Washington, DC: Carnegie Institution Publication No. 164.
- Hall R A Jr. (1966). *Pidgin and creole languages*. Ithaca, NY: Cornell University Press.
- Holm J (1989). *Pidgins and creoles* (2 vols). Cambridge: Cambridge University Press.

- Keesing R (1988). *Melanesian Pidgin and the Oceanic substrate*. Stanford: Stanford University Press.
- Lefebvre C (1998). *Creole genesis and the acquisition of grammar: the case of Haitian Creole*. Cambridge: Cambridge University Press.
- McWhorter J (1998). 'Identifying the creole proto-type: vindicating a typological class.' *Language* 74, 788–818.
- McWhorter J (2001). 'The world's simplest grammars are creole grammars.' *Linguistic Typology* 5(2), 125–166.
- Mufwene S S (1986). 'Les langues créoles peuvent-elles être définies sans allusion à leur histoire?' *Etudes Créoles* 9, 135–150.
- Muysken P (1988). 'Are creoles a special type of language?' In Newmeyer F J (ed.) *Linguistics: the Cambridge survey 2: Linguistic theory: extensions and implications*. Cambridge: Cambridge University Press. 285–301.
- Roberts S J (2000). 'Nativization and the genesis of Hawaiian Creole.' In McWhorter J (ed.) *Language change and language contact in pidgins and creoles*. Amsterdam: John Benjamins. 257–300.
- Romaine S (1988). *Pidgin and creole languages*. London: Longman.
- Romaine S (1992). *Language, education and development: urban and rural Tok Pisin in Papua New Guinea*. Oxford: Oxford University Press.
- Singler J V (ed.) (1990). *Pidgin and creole tense-mood-aspect systems*. Amsterdam: John Benjamins.
- Thomason S G & Kaufman K (1988). *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Velupillai V (2003). *Hawaii Creole English: a typological analysis of the tense-mood-aspect system*. Basingstoke: Palgrave Macmillan.

Pidgins and Creoles, Variation in

G Escure, University of Minnesota, Minneapolis, MN, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction: Creole Myths

Pidgins and creoles have long been characterized as ungrammatical and their speakers as uneducated. This bias is illustrated in the following excerpt from the first novel completely written in a French-based creole (Guyanais), a stinging satire of French colonial society in Cayenne offered through the voices of two Creole characters: Atipa, a gold miner, and his friend Bosobio:

- (1) **Atipa:** Nu kriol pa gen reg ku franse
We creole not have rule like French
- (2) nu sa pale li ku nu wle...
we TOP speak it as we want
- (3) gremesi bunge landan nu lang
Thank god in we language
- (4) nu pa benzwen okjupe di sintas...
we not need worry of syntax
- (5) Mo rin save sintas-la sa lang ye
me nothing know syntax-DET TOP language that
- (6) ka pale la konsey ke la tribinal
IMPERF speak DET council like DET tribunal
- (7) **Bosobio:** a pu sa li gen un ta di zafe
TOP for that it have one lot of business
- (8) mo pu ka konpren ni la
me not IMPERF understand neither at
tribinal ni la fomasi-la
tribunal nor DET pharmacy-DET

Atipa: 'We Creoles do not have grammatical rules as in French, we speak just as we like. Thanks to God who gave us our language, we don't have to worry about syntax. I don't know anything about syntax, it's the language they use at council meetings, and at the tribunal. **Bosobio:** That's why there are so many things I don't understand, either at the tribunal, or the pharmacy.' *Atipa* (Parépou, 1885)

Atipa's anonymous author, who used the pseudonym of Alfred Parépou, neatly summarizes the myths attached to creoles, and their social correlates: creoles are not real languages ('we speak as we want'; 'creole has no syntax'); furthermore, creole speakers are excluded from official business and basic social services. Yet, the author demonstrates that this nonlanguage can be used to write a 227-page novel!

The young languages we call pidgins and creoles are universally engendered in the context of traumatic situations such as slavery, indenture, or migration. Although pidgins and creoles differ in the scope of their social functions – pidgins are short-term creative attempts at producing *lingua francas*, whereas creoles are native vernaculars – they have in common that they are oral languages spoken by marginalized groups, are rarely acknowledged as valid grammatical systems, even by their own speakers, and are therefore rarely written. *Atipa* is a major exception, but even now literature fully written in creole is scarce.

This article identifies some of the linguistic conflicts and choices that face pidgin and creole speakers in their social networks. Rather than providing an overview of the wide range of variation that occurs

in creole communities around the world, I will focus on a few representative examples.

Variation in Pidgins

Pidgins are generally short-term varieties restricted to specific social domains or occasional events such as seasonal trade activities. Pidginization has often been defined as ‘imperfect’ acquisition of the target language, but this characterization is debatable. The objective in any of the emergency situations that give birth to new varieties is basic communication, not native-like fluency in the dominant language. If one accepts this pragmatic goal as a realistic option, it is clear that linguistic variability must have been present from the very beginning of the contact.

Since the rapid production of an operational lingua franca is crucial, and happens without the benefit of proper instruction, pidgin development can be expected to be highly variable. Some of the strategies widely used in pidginization are illustrated in the following sample of CPE (Chinese Pidgin English), a lingua franca that developed in the 19th century as British ships traded in Canton, and Cantonese (Yue)-speaking Chinese (Chinese, Yue) merchants and servants made the effort to communicate in English with Europeans. CPE evidence is represented in a large number of occasional (and not necessarily accurate) observations made by Europeans. CPE combines English lexicon and Chinese substratal influences, such as paratactic structures rather than subordinating syntax, the use of elements such as *suppose* to separate propositions, and of classifiers such as *piece* before nouns. Some of these features occur widely in pidgins (and creoles), but others do not, and are thus traceable to transfer from Chinese, such as the usage of a classifier in (15). Variation is illustrated below in sentences excerpted from a large unpublished corpus made available by Philip Baker (CPE Corpus, 2004). The pidgin sometimes functions as a pro-drop language (absence of subject pronouns in [9–10, 13]), but sometimes not, using indiscriminately subjective or objective pronouns, since Chinese has no case marking (2004: 11–12) [translation is provided only when the meaning may be unclear]:

- (9) This have every poor place, and very poor people;
no got cloaths, no got rice, no got hog; no got
nothing; only yam, little fish, and cocoa-nut;
no got nothing make trade, very little make eat.
- (10) No got fowls, have got chicken [...] no can tell,
must first make weigh.
- (11) Me think have go Pekin.
- (12) Suppose he have no got eye, how can him see?
Suppose he no can se, how can him walkie?

- (13) Suppose cheat a little can do, suppose cheat too
muchy no can.
- (14) Suppose no gib lice, how can lib?
‘If you don’t give me rice, how can I live?’
- (15) One piece man [...] How much piece masts hab
got you ship, how many piece guns, shot and
powder? How much piece woman, cow
childes and bull childes?
‘One man [...] How many masts have you got
on the ship, how many guns, bullets and
powder? How many heifers and calves?’

Variation in Creoles

Since creoles are more numerous and better documented than pidgins – but note that many contemporary creoles are called ‘Pidgins,’ such as Nigerian Pidgin, or Tok Pisin – the remainder of the article will discuss two related issues that have lately dominated the field of creole studies.

First, the reality and structure of the ‘creole continuum’ is examined. Creoles (like pidgins) were never isolated from their lexifiers. The social background of native creole vernaculars was such that their subaltern speakers were always in contact with the language of dominant social strata, but in differential ways. Some individuals (i.e., house slaves) had better access than others (field slaves) to the target language (TL), which may have been either socially or demographically dominant. Moreover, the available TL was not necessarily the standard (or prestigious) version of the lexifier: it may have been a nonstandard variety of the European language(s), for example, in contacts between slaves and overseers or ship hands, and thus learners of different varieties were likely to interact and use different versions of the TL as *lingua franca*. In addition, demographics (such as relative proportion of European speakers of the TL and Africans in contact) determined the outcome of the creolization process during the formative period (Chaudenson, 1992). The proportions of speakers varied according to the region or the household, which explains the linguistic differences between neighboring varieties – e.g., between Morisyen (in Mauritius) and Réunionnais (Réunion Creole French; in Réunion), both in French-colonized islands in the Indian Ocean; or between Jamaican Creole and Bajan (Barbados), both spoken in English-colonized Caribbean islands. In those two parallel cases, whites outnumbered slaves in Réunion and in Barbados, but the opposite was true in Mauritius and Jamaica. Consequently, Morisyen and Jamaican are more ‘creolized’ than their counterparts. This designation means that the most basilectal varieties in Mauritius and Jamaica have no equivalents in Réunion and Barbados:

Bajan and Réunionnais have more restricted repertoires, ranging only between mesolects and acrolects.

Linguistic variability is to be expected at every stage of the language history. Most previous colonies remained economically dependent on European (or other) nations, even after independence was granted. Because of the continuing contiguity of prestigious and stigmatized varieties – greatly facilitated by the greater availability of education – language stabilization is counterintuitive in any creole context, which does not exclude the existence of a regular creole system. Similarly, single-style speakers are rare, even in remote rural areas. However, some varieties called ‘radical creoles’ (Saramaccan for example) are assumed to be somewhat stable, restricted to conservative varieties, and not associated with a continuum. This situation may be the consequence of group isolation, as suggested by Atipa in the Guyanais quotation shown above, but it is doubtful that such social contexts still exist. With some rare exceptions, the concept of the creole continuum effectively captures the flexible reality of contact vernaculars.

Secondly, the issue of ‘decreolization’ – that is convergence toward the dominant language, and concomitant loss of the creole – is re-evaluated. Although pidgins generally disappear, or evolve into more complex varieties, many creoles thrive and retain high covert prestige in their native communities, even as they interact with dominant or official languages.

The Creole Continuum

Since creoles are still overwhelmingly considered by public opinion to be corruptions or distortions of official languages, speakers of those marginalized varieties are bound to acquire some version of the local standard. Literacy is widely available now, and the ‘proper’ medium of instruction is naturally the official language (e.g., English in Belize; French in Martinique, Portuguese in Cape Verde, etc.). However, the standard model is rarely present in the classroom, as local teachers have variously acquired their own version of the standard, thus contributing to the continued linguistic variability observable in creole areas.

Early pioneering studies viewed creoles as static nonstandard approximations of their lexifiers. This perspective implied that creole speakers consistently used a predictable nonstandard system. But a few innovative analyses of creole variation led the way to a more realistic understanding of linguistic repertoires. DeCamp (1971) in his description of Jamaican Creole was the first to use the concept of continuum as an analytical tool in complex linguistic situation. He referred to a wide range of linguistic options that were available to the creole speaker, as illustrated

in variants such as: *mi tel am/a tel im/a told him*, pointing out the lack of clear separation between variants, and the myth that there are only two varieties of language.

Pidgin and creole speakers are constantly fluctuating between two poles, the vernacular, which is appropriate in familiar, at home and in group situations, and the formal standard, which is required in official contexts, typically work and out group situations. But speakers’ repertoires are not restricted to two clearly bounded varieties; they spread over a continuum of overlapping forms, whose specific representations are dictated by the social, ethnic, or gender contexts, the competence and adaptability of individual speakers, and other psychological factors. The ‘creole continuum’ aptly captures the absence of any clear boundary separating the various speech types available within any Creole community.

This continuum can be divided into three broad variety groupings (or ‘lects’): ‘basilects,’ ‘mesolects,’ and ‘acrolects.’ Basilects are the most vernacular varieties that linguists have typically described as creoles. Acrolects are often used to refer to Creole speakers’ production of the local standard language, yet they are not identical to that standard; they are usually L2 versions of the standard. Finally, mesolects are located somewhere between basilects or mesolects, yet are not imperfect approximations of the acrolect. Mesolects have their own structure and their own *raison d’être*.

Bickerton (1975) was probably the first to complete a comprehensive analysis of the language spectrum for Guyanese Creole (English-based creole), and his novel approach stimulated a number of creole studies that adopted the concepts of continuum, and the related notion of implicational scales, as analytical devices. To cite just a few studies of English-based creoles: Washabaugh (1975), Herzfeld (1978), Craig (1980), Escure (1981), Singler (1984), Rickford (1987), Crowley (1990), Patrick (1992), Aceto (1996), Smith (2002). Studies of French-based varieties include Ludwig (1989), Chaudenson (1992), Lefebvre (1998), Corne (1999), and many more. Some examples of variability across creole continua are provided below, illustrating variation in lexical semantics, phonology, and morphosyntax in samples taken from two English-based creoles, Ghanaian Pidgin English (West Africa) and Belizean Creole (Central America).

Lexical Semantics

The naming of body parts offers a well-known example of semantic differentiation at the word level. Many creoles display substrate influences in the

naming of limbs, with the transfer of African semantic structures into Indo-European lexicon: thus, following Bantu and Kwa practice of using one single word to refer to the whole limb, English-based creoles (Belizean, Jamaican) use *fut* to refer to both 'foot' and 'leg', and *han* to refer to 'hand' and 'arm' (but Nigerian Pidgin used *leg* for both 'foot' and 'leg', though it uses *han* as the generic upper limb term). In Portuguese, creoles such as São Tomé, the equivalent Portuguese words are used with the same semantic range. The same feature occurs in Bislama (also English-based, spoken in Vanuatu), though the substratal influence is Austronesian in this case. When speakers of those creoles switch to acrolects, they then use the appropriate term. For example, a Creole boy (in Belize) said (showing his calf): *Wan shaak bait mi fut hia*, 'A shark bit my leg here,' but in the next minute, he switched to an acrolect: *Main da maskito pan yu leg*, 'Mind that mosquito on your leg' (Escure, 1990).

Education and Llectal Level (Ghanaian Pidgin)

The short dialogue shown below, taken from a radio commercial in Accra, Ghana, illustrates particularly well subjective attitudes toward the varieties available to creole speakers: the creole (Ghanaian Pidgin) is attributed to the uneducated speaker (taxi driver), while the engineer speaks Standard Ghanaian English (acrolect). The transcription represents basilectal features in the driver's speech: phonological features (absence of interdentals, absence of postvocalic /r/, and of /h/), morphosyntactic features (use of preverbal imperfective *de*, unmarked past, relativizer *we*, single preverbal negative element). On the other hand, the engineer uses 'flawless' English grammar (but Huber's audio version reveals acrolectal phonetic variants):

(16) **Driver:** ju sabi ma padi adzeman, i de draiv tata
bos we in masta bai fo hia
'You know my friend Agyeman, he
drives a Tata bus that his master
bought here'

(17) **Engineer:** The Yellow Cab Company Ltd?

(18) **Driver:** jes, i no de bring am fo sevisin en
mentenans, en i de poches in spepas
fo evriwea.

'Yes, he doesn't bring it here for
servicing and maintenance, and he
buys his spare parts from everywhere'

(19) **Engineer:** Is this Tata vehicle on the road?

(20) **Driver:** No, i de brok daun plenti-plenti.
'No, it keeps breaking down. (Huber,
1999: 271)

Llectal Variation (Belizean Creole)

A few texts drawn from an unpublished Belizean corpus by Escure (1990) illustrate the extensive range of the creole continuum, starting with the most extreme lects, basilects and acrolects, then addressing the elusive mesolect.

Basilect (Nansi Story)

Miss Dolly (a 60-year old woman from Placencia) tells a traditional tale (Nansi story). This story evidences some prominent basilectal features:

- The use of the preverbal aspectual morpheme *de* (e.g., *everibadi de dans* 'everybody dances/keeps dancing') is best defined as an imperfective, as it may have progressive and habitual/iterative functions.
- The nonmarking of past (e.g., *di dans stat tu brok op* 'the dance started to break up.')
- The creole reinterpretation of some old preterites as bare verbs (e.g., *brok* 'break').
- The occurrence of a different preverbal past morpheme *me* (sometimes with anterior value), which helps distinguish between two sequential past events (*di mjusik me de ple* 'the music was playing' as background event to the crowd leaving the dance-hall). Here, the past morpheme is also combined with the imperfective marker indicating continuing action (see Escure, 2004 for a more complete list of basilectal features).

(21) Dis da wan taim nou dei had
This TOP a time now they had
wan dans evribadi de dans
a dance everybody IMPERF dance
'Once upon a time, they had a dance, everybody would dance'

(22) bra taiga bra dag bra everibadi
Brother Tiger Brother Dog Brother Everybody
dans Evribadi de dans
so Evrybody IMPERF dance
'Brother Tiger, Brother Dog, Brother Everybody, so
everybody would dance'

(23) buldag de dans kou de
bull dog IMPERF dance cow IMPERF
dans evribadi
dance everybody
'the bulldog dances, the cow dances, everybody'

(24) big pati de gown tuwad midnait nou
big party IMPERF go.on towards midnight now
di dans stat tu brokop
the dance start to break.up
'it's a big party, towards midnight the dance ended'

(25) bika wan fait stat evribadi stat tu fait
because a fight start everybody start to fight
'because a fight started, everybody started to fight'

- (26) evribadi de tekdu dem bati
everybody IMPERF take down DET butt
 an de kot
and IMPERF cut
 'and everybody started to go and they left'
- (27) an dat waz di en a di pati
and that was the end of the party
 'and that was the end of the party'
- (28) bot di mjuzik we me de plei
but the music that ANT IMPERF play
 'but the music that was playing'
- (29) i go laik dis: zinzinzin. vajalin
it go like this: zinzinzin. violin
 da me di mjuzik
TOP IMPERF the music
 'it went like that: *zinzinzin* it's the violin that
 was playing' (Nansi story, Escure, 1990)

An additional example shows how creole marks irrealis modality (unrealized events) through the combination of the anterior marker *me* and the future marker *wan* – a grammaticalized form of the verb 'want':

- (30) R. wan tek wan korespondens kos.
R. FUT take a correspondence course
 'R. will take a correspondence course.'
- (31) i me wan tek it befo i kum awt.
he ANT FUT take it before he come out
 'He would have completed it before he graduates.'
- (32) i me de plan fu tek it
he ANT IMPERF plan to take it
 'He was planning to take it'
- (33) da di taim de tem don di kos don.
that the time the term done the course done
 'so that by the time the term is over, his course would have been completed.' (Escure, 1990)

Acrolect

The acrolect is a local standard that differs from external standards. Since acrolects are typically the result of late acquisition, probably through school education, inconsistencies are most likely to occur at this lectal level, depending on social factors, such as an individual's relative access to the standard, or psychological factors, such as the speaker's identity and intent to converge toward the standard. The acrolect generally differs phonetically from its lexifier (in the case of Belizean English, it differs from RP-British English). Most common distinctions include the systematic or occasional absence of interdental fricatives, and variation in vowels (for example, lack of distinction between tense and lax vowels). Acrolects generally use standard grammar and morphology, for

example, past verbs are now marked, preverbal morphemes are absent, the copula *be* is introduced, and so forth, but more variation occurs in upper mesolects, that vague area between the widely used labels of 'English' and 'broken English.' Thus, nonstandard morphological features may be part of an acrolectal version (for example, absence of copula/auxiliary; lack of 3SG agreement; hypercorrect past inflection, or pronoun variation), as are pragmatic mechanisms (such as the fronting of topics). The following sentence displays both *be* presence (*dei were expectin*) and absence (*would willin*):

- (34) Dei we espektin samwan den wu
They were expecting someone then who
 wud wilin tu tekop amz
would willing to take up arms
 'They were expecting someone who would be willing to take up arms.' (Escure, 1990)

Newspapers often exhibit similar linguistic features, whether unwittingly or as intended for special effect:

- (35) I can **recalled** a very shocking incident [...] One may come to the conclusion that an abundance of ignorance **exist** within [...] This area has long been mean, but never **have** it been so lethal [...] Such an attitude gathers strength from its own **existent**, the longer it **persist**, the deeper it roots grow. ('Help our troubled & lost generation' *Alkebulan* (Belize), January 21, 1994: 2)

But an article on local politics – discussing the rival political party (PUP) – inserts some basilectal phrases in the middle of a standard text for emphasis (here sarcasm, shown in bold characters in the original text):

- (36) [...] their plaintive wail when all else fails is victimization, translation: **A fri'ten bad**. [literally, I frightened bad 'I am afraid']
- (37) [...] Houses are being built [...] **And would you believe it, the PUP bex bout that**. [literally, PUP vexed about that 'the PUP is annoyed about that'] ('The observer' *The People's Pulse* (Belize), April 17, 1994: 14)

Mesolect (The Village Midwife)

Mesolects can be defined as intermediate varieties, but they are not mere approximations of the standard: they have their own internal motivation and place in the social life of continuum users. Individuals who control the whole range of the continuum select the mesolect in well-defined situations – when addressing an older person, or the members of another ethnic group, or dealing with a serious topic. There are issues of respect, of formality, and of identity

involved in such choices, so it is not possible to speak of 'basilectal' or 'mesolectal' speakers, except to say that in context A, an individual is a basilectal speaker, but in context B, the same speaker is a mesolectal speaker.

In the following excerpt, Miss Dora, a 75-year old midwife who has delivered all the village babies for the last 50 years, uses neither a basilect nor an acrolect. She has native competence in the creole vernacular, but selects the mesolect when recounting her professional activities with her nephew. Characteristics of this mesolect include absence of copula, unmarked past, and an occasional preterite form (*had*) as well as the auxiliary *don't* (instead of simple preverbal negative). Mesolects generally imply avoidance of basilectal morphemes, but this implication is not always the case: at crucial peaks of the narrative, Miss Dora uses the TMA creole morphemes *de* (Imperfective) *me* (Past Anterior), as well as the expression *don ded* 'completely dead', a common use of the perfective marker 'done' to emphasize the finality of death. Note also the creole use of *lef* for 'leave' (*I had to lef dat* 'I had to leave/stop that'), one of a few verbs whose neutral form is a relexified irregular preterite.

(38) Da sem taim tu peshen kum in
At same time two patient come in

(39) wan mada da di ilevent bebi im gat
a mother TOP the eleventh baby she got
en di ada wan da di naint
and the other one TOP the ninth

(40) en de riali nat sapoz tu got bebi da vilidg
and they really not supposed to get baby at village

(41) bot den dei don wan go da haspital [...]
but then they don't want go to hospital

(42) wel a had a fait wid di bebi
well I had a fight with the baby
bika di bebi hed kum
because the baby head come

(43) bot di ada paat a di bodi wont kum [...]
but the other part of the body won't come

(44) den di aftabat kyan kum
then the afterbirth can't come

(45) a had tu lef dat wan an di ada wan redi
I had to leave that one and the other one ready

(46) a swab shi af [...]
I swab her off

(47) an den shi lef wika stil
and then she stay weaker still
de hemoredg
IMPERF hemorrhage

(48) an wen a give shi dat an fainali i kwait dawn
and when I give her that and finally she quiet down

(49) an shi an mai sista me tink di
and she and my sister ANT think the
bebi don ded
baby PERF dead

(50) a do mawt tu mawt bridin an
I do mouth to mouth breathing and
di bebi big bwai naw.
the baby big boy now

'Two patients came in at the same time. One mother was delivering her eleventh baby, and the other her ninth [...]. They are not really supposed to deliver in the village, but they don't want to go to the hospital [...]. Well I had to struggle with the (first) baby because its head was coming out, but not the rest of its body [...] then the after-birth wouldn't come [...] I had to leave that one (first mother) to go to the other one who was ready (to deliver) [...] I cleaned her (second mother) up [...] (first mother) remained weak, and was still hemorrhaging [...] When I gave her (herbs) she (first mother) finally settled down. She and my sister thought that the baby was already dead [...] But I did mouth-to-mouth resuscitation, and the baby is now a big boy.' (The Village Midwife, Escure, 1990)

Decreolization

Schuchardt's 'life cycle' concept (1883) became DeCamp's 'postcreole' continuum (1971). This developmental hypothesis suggests that creoles eventually merge with the standard, assuming that the continuum is the result of decreolization (loss of the creole). However, the data presented above suggest that the acquisition of acrolects or near-standard varieties – obviously facilitated by access to education and standard speakers after emancipation – does not necessarily entail concomitant loss of basilectal segments. Individuals, with few exceptions, are generally found to control a wide repertoire. Empirical studies show that they don't lose their native variety just because they have acquired a new one – no more than L2 acquisition would entail loss of L1, except in extreme situations leading to language death.

The ability to handle alternate codes has been explained in terms of the 'dual standard,' or the 'covert' vs. 'overt prestige' dichotomy: as subaltern groups gain access to education, they become increasingly motivated or obligated to learn the standard as a means of improving their social position. Creole values may thus be overtly despised but secretly respected, whereas the values of the high-status group are overtly respected and secretly despised. As is the case in any multilingual context, individuals make linguistic choices that reflect their allegiance or close associations with either the dominant social group (usually speaking standard varieties), or

the peer group, or both. The 'linguistic market' sociological model of linguistic production and expression also captures the relation between linguistic system (*l'habitus linguistique*) and linguistic market (*le marché linguistique*) (Bourdieu, 1982).

Such perceptual differences still mirror the historical colonial bias and the shift to a new social order. They also explain why creole languages offer such a wide range of linguistic possibilities. The linguistic spectrum captures the multiple nuances required in various human contact situations. The very nature of its flexibility ensures that **all** varieties remain active and operational, and contradicts the view that there is an ineluctable move toward the standard, since native (basilectal) values are highly prized, though covertly.

According to this perspective, decreolization is not diachronic change (although regular change naturally occurs), but rather repertoire extension and code switching. There is no postcreole continuum if the creole is still vigorous, as in Belize, or Haiti, or Papua New Guinea. There is a postcreole situation when the creole has lost most of its speakers, as in Louisiana, in which the confusion of the French-based creole with Cajun (a French Canadian dialect), the import of French teachers from metropolitan France, the dominance of English, and generally the low status of black speakers have probably contributed to the receding state of Louisiana Creole.

Conclusion

The field of creolistics has expanded considerably as new sociohistorical sources have redefined our understanding of the early stages of language genesis and development, and as more empirical field studies have offered testing grounds for theoretical and sociolinguistic models of language use and language development. Subfields of linguistics (historical linguistics, sociolinguistics, and theoretical linguistics more specifically) can benefit from the current state of knowledge in pidgins and creoles. New creoles encapsulate the linguistic effects of the violent social history that most of humanity has been subjected to. Language development is closely dependent on the economic and political features of the societies in whose context they emerged, and current linguistic variability serves to illustrate further the correlation that exists between linguistic structures and social aspects. Creole speakers use polylectal systems, rather than monolithic grammars, and this aspect should be highly relevant to theoretical models that focus on abstract generalizations but overlook the human language ability to juggle multiple systems.

Bibliography

- Aceto M (1996). 'Variation in a variety of Panamanian Creole English.' Ph.D. Diss.: Austin: University of Texas.
- Bickerton D (1975). *Dynamics of a creole system*. Cambridge: Cambridge University Press.
- Bourdieu P (1982). *Ce que parler veut dire*. Paris: Fayard.
- Chaudenson R (1992). *Des îles, des hommes, des langues: essais. sur la créolisation linguistique et culturelle*. Paris: L'Harmattan.
- Corne C (1999). *From French to Creole: the development of new vernaculars in the French colonial world*. London: University of Westminster Press.
- CPE Corpus (2004). Chinese Pidgin English corpus. (Unpublished).
- Craig D (1980). 'A creole English continuum and the theory of grammar.' In Day R (ed.) *Issues in English creoles*. Heidelberg: Julius Groos Verlag. 111–132.
- Crowley T (1990). *From Beach-la-mar to Bislama: The emergence of a national language*. Oxford: Oxford University Press.
- DeCamp D (1971). 'Towards a generative analysis of a post-creole creole continuum.' In Hymes D (ed.) *Pidginization and creolization of languages*. Cambridge: Cambridge University Press. 349–370.
- Escure G (1981). 'Decreolization in a creole continuum: Belize.' In Highfield A & Valdman A (eds.) *Historicity and variation in creole studies*. Ann Arbor: Karoma Press. 27–49.
- Escure G (1990). Corpus of Belizean varieties. (Unpublished).
- Escure G (2004). 'Belize and other Central American varieties: morphology and syntax.' In Schneider E (ed.) *A handbook of varieties of English: The Americas and the Caribbean*. The Hague: Mouton. 315–342.
- Herzfeld A (1978). 'Tense and aspect in Limón Creole: a sociolinguistic view towards a creole continuum.' Ph.D. Diss. Department of Linguistics, University of Kansas.
- Huber M (1999). *Ghanaian Pidgin English in its West African context*. Amsterdam & Philadelphia: John Benjamins.
- Lefebvre C (1998). *Creole genesis and the acquisition of grammar: the case of Haitian Creole*. Cambridge: Cambridge University Press.
- Ludwig R (ed.) (1989). *Les créoles français entre l'oral et l'écrit*. Tübingen: Gunther Narr Verlag.
- Parépou A (1885). *Atipa*. (Ré-édité 1980). Paris: Editions Caribéennes, Collection UNESCO d'oeuvres représentatives. (Série créole.)
- Patrick P (1992). *Urban Jamaican Creole*. Amsterdam & Philadelphia: John Benjamins.
- Rickford J (1987). *Dimensions of a creole continuum*. Stanford: Stanford University Press.
- Schuchardt H (1883). *Kreolische studien v. Über das Malaiospanische der Philippinen*. Vienna: Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien.
- Singler J (1984). 'Variation in tense-aspect-modality in Liberian English.' Ph.D. Diss. Los Angeles: University of California.

- Smith G (2002). *Growing up with Tok Pisin: contact, creolisation and change in Papua New Guinea*. London: Battlebridge.
- Washabaugh W (1975). 'Variability in decreolization on Providence Island, Colombia.' Ph.D. Diss. Chicago: Wayne State University.

- (1994). 'Help our troubled & lost generation.' *Alkebulan*. (Belize), January 21, 2.
- (1994). 'The observer.' *The People's Pulse*. (Belize), April 17, 14.

Pitjantjatjara/Yankunytjatjara

C Goddard, University of New England, Armidale, Australia

© 2006 Elsevier Ltd. All rights reserved.

Pitjantjatjara, along with its neighboring dialect Yankunytjatjara, are part of the Western Desert Language (WDL) – a vast dialect continuum located in the arid and sparsely populated central and western inland of Australia (*see Australian Languages; Australia: Language Situation*). The two dialects are commonly referred to jointly as P/Y. There are about 2500 P/Y speakers. Since it is still being acquired by children, P/Y counts as one of the less endangered of Australian languages.

As a typical Pama-Nyungan language, WDL is agglutinative (chiefly suffixing), with well developed systems of nominal and verbal inflection. Canonical constituent order is S (O) (PP) V, but can vary rather freely. Ellipsis of third person arguments, when the referent can be understood in context, is common. Nominal inflection is of the split ergative variety. The verbal system has eight tense-aspect-mood categories, with complex allomorphy governed by a system of four conjugational classes. Serial verb constructions of several types, abound. P/Y has a switch-reference system in several subordinate clause types, and in coordinate constructions. Aspects of P/Y have been studied by a number of linguists. There are three major grammars and a substantial dictionary, a range of pedagogical material, and a variety of specialized linguistic studies. A wide range of vernacular texts of traditional stories, ethnoscience, and oral histories has been published locally.

Sociocultural and Historical Aspects

The similarity between the two dialects has been reinforced by shared historical experiences in the wake of European intrusion early last century, including long periods of co-residence on mission and government settlements, and, subsequently, in self-managed Aboriginal communities. Most speakers

now co-reside on Aboriginal-owned lands in the north-west of the state of South Australia and adjacent areas in the Northern Territory. For a long time Pitjantjatjara was the prestige variety, because it had been adopted by missionaries at Ernabella for Bible translation and use in Christian worship, and was subsequently used in bilingual education programs in local primary schools. Lately the two dialects have been moving towards parity of esteem. Many Australian and international tourists encounter P/Y when they visit the Uluru National Park in central Australia.

Though the two dialects share about 80% vocabulary, there are a number of prominent dialect-specific words, and these form the basis of the traditional WDL system for referring to speech varieties. Pitjantja-tjara and Yankunytja-tjara are based on alternative (nominalized) forms of verbs meaning 'come/go' (suffix *-tjara* means 'having'). Northern and southern varieties of Yankunytjatjara can be termed Mulatjara and Matutjara, respectively, based on alternative forms of the adverb 'true.' In earlier times, there was a multiplicity of such terms in use. The system was highly relativistic, allowing for cross-cutting categorization and for different levels of inclusiveness, which suited the traditional mobile and dynamic social economy. These days the terms Pitjantjatjara and Yankunytjatjara have acquired more stable and 'name-like' sociopolitical functions.

Traditional P/Y culture is replete with symbolism (totemism) and religious myth. There are hundreds of Dreaming stories, songs, and ceremonies. There is a large body of traditional folktales for children. Many P/Y speech practices have parallels in the other languages of Australia. These include the existence of hortatory rhetoric (*alpiri*), elaborate verbal indirectness practiced with certain categories of kin (and total avoidance with others), and prescribed 'joking relationships' characterized by mock insult and abuse. There is a taboo against using the names of recently deceased persons in the presence of bereaved relatives. An auxiliary register, i.e., a special vocabulary (termed *anitji*), is used during ceremonial times.

Structure

Phonology

P/Y has 17 consonant phonemes: see Table 1. There are five places of articulation, each with a stop and a nasal. There are two series of apicals, i.e., consonants pronounced with the tongue tip as active articulator: alveolar and post alveolar (retroflex). There is a single laminal series, with the tongue blade as active articulator. There are three vowels (*a*, *i*, *u*), each with a length distinction, though long vowels are not common and are confined to initial syllables.

P/Y phonotactics stipulate that a word must have at least two vowels, with long vowels counting as two for this purpose. Several morphophonemic rules refer to whether a stem has an odd or even number of vowels, making it convenient to work in terms of morae, with long vowels counting as two morae. Words usually start with a single consonant and never with more than one. Inside a word, CC clusters occur subject to strict limitations. Most common are homorganic ‘nasal/lateral + stop’ sequences. Only a very limited set of consonants (*n*, *ny*, *n̩*, *l̩*, *r*) is permitted word-finally, and then only in the Yankunytjatjara variety. In Pitjantjatjara, consonant-final words are blocked by addition of the syllable-*pa*.

Morphology and Syntax

A number of these features are illustrated in the text extract at the end of this section.

Nominal Morphology

The case system includes nominative, ergative, accusative, genitive/purposive, locative, allative, ablative, and perlative cases. Typically a case-marker is applied only to the final word of an NP. Since modifiers generally follow their heads, a typical multi-word NP looks like: *wati pulka kutjara-ku* [man big two-PURP] ‘for two big men.’ Like most other Pama-Nyungan languages, there is a split marking system for the core cases. For both nouns and pronouns, the

nominative case is unmarked. With nouns, accusative case goes unmarked but there is a marked ergative form (with *-ngku/-lu* or a variant). With pronouns, the ergative goes unmarked but there is a marked accusative (with *-nya*). Split case-marking is sometimes described in terms of two distinct case systems: nominative-accusative for pronouns and ergative-absolutive for nouns. Aside from being less economical, such an analysis has difficulty with various complex NP constructions involving both nouns and pronouns. For example, inalienable possession constructions can bring body-parts and pronouns into a single NP, and inclusive constructions can bring names and pronouns into a single NP. For example, to say that someone hit me on the head, one uses the NP *ngayu-nya kata* [1SG-ACC head:ACC] ‘me head.’ To say that Kunmanara and someone else did something to someone, one uses the NP *Kunmanara-lu pula* [name-ERG 3DL:ERG].

Ergative and locative case allomorphy depends on whether the word to be marked is vowel- or consonant-final, and on whether the NP is an ordinary noun-phrase, on the one hand, or a pronoun or proper noun, on the other. Ergative is *-ngku* (common) or *-lu* (proper) with vowel-final words, and otherwise *-Tu* (where *T* is a homorganic stop). Locative is *-ngka* (common) or *-la* (proper) with vowel-final words, and otherwise *-Ta*. Genitive/purposive case is marked with *-ku* (nouns) or *-mpa* (pronouns, except for 1SG *ngayu-ku*). Locative also expresses instrumental and comitative functions; e.g., *punu-ngka* [stick-LOC] ‘with a stick,’ *untal-ta* [daughter-LOC] ‘with (my) daughter.’

Pronouns distinguish singular, dual, and plural numbers (see Table 2). Most WDL dialects also have enclitic or ‘bound’ pronouns that can be used instead of or in addition to free pronouns. They appear attached to the first phrase of a sentence, conjunctions counting as phrases for this purpose. P/Y has the following defective set – nominative/ergative: *-na* 1SG, *-n* 2SG, *-li* 1DU, *-la* 1PL, *-ya* 3PL; accusative: *-ni/-tja* 1SG, *-nta* 2SG, *-linya* 1DU, *-lanya* 1PL. Bound pronouns are not obligatory in P/Y, though they are common. There are four demonstrative stems: *nyanga* ‘this,’ *pala* ‘that,’ *nyara* ‘that over there,’ and the anaphoric demonstrative *panya* ‘that one, you know which.’

Verbs

All WDL dialects share a similar system of tense-aspect-mood categories and four conjugational classes, though the details differ from dialect to dialect. The P/Y categories are: present, past, past

Table 1 Pitjantjatjara/Yankunytjatjara consonant phonemes, in standard orthography (Goddard, 1985: 11)

| | Apical | | Laminal | | |
|----------|----------|--------------|---------|----------|--------|
| | Alveolar | Postalveolar | Dental | Bilabial | Dorsal |
| Stops | t | t̩ | tj | p | k |
| Nasals | n | n̩ | ny | m | ng |
| Laterals | l | l̩ | ly | | |
| Tap | r | | | | |
| Glides | | r̩ | y | | w |

Table 2 Pitjantjatjara/Yankunytjatjara subject free pronouns (Goddard, 1996: xi)

| Subject | Singular (sg) | | Dual (du) | | Plural (pl) | |
|---------------|-------------------------------|---------------|----------------|------------|----------------|--------|
| First person | <i>ngayu(lu)</i> ^a | 'I' | <i>ngali</i> | 'we two' | <i>nganana</i> | 'we' |
| Second person | <i>nyuntu</i> | 'you' | <i>nyupali</i> | 'you two' | <i>nyura</i> | 'you' |
| Third person | <i>palu(ru)</i> | 'he, she, it' | <i>pula</i> | 'they two' | <i>tjana</i> | 'they' |

^a The syllables in parentheses are dropped when case suffixes are added.

Table 3 Pitjantjatjara/Yankunytjatjara verbs (Goddard, 1985: 90)

| | (Ø) 'talk' | (l) 'bite' | (ng) 'hit' | (n) 'put' |
|---------------------------|--------------------|-------------------|---------------------|---------------------|
| Imperative | <i>wangka</i> | <i>patjala</i> | <i>puwa</i> | <i>tjura</i> |
| Past (perfective) | <i>wangkangu</i> | <i>patjanu</i> | <i>pungu</i> | <i>tjunu</i> |
| Imperative (imperfective) | <i>wangkama</i> | <i>patjanma</i> | <i>pungama</i> | <i>tjunama</i> |
| Present (imperfective) | <i>wangkanyi</i> | <i>patjani</i> | <i>punganyi</i> | <i>tjunanyi</i> |
| Past (imperfective) | <i>wangkangi</i> | <i>patjaningi</i> | <i>pungangi</i> | <i>tjunangi</i> |
| Future | <i>wangkaku</i> | <i>patjalku</i> | <i>pungkuku</i> | <i>tjunkuku</i> |
| Characteristic | <i>wangkapai</i> | <i>patjalpai</i> | <i>pungkupai</i> | <i>tjunkupai</i> |
| Serial form | <i>wangkara</i> | <i>patjara</i> | <i>pungkula</i> | <i>tjunkula</i> |
| Nominalized form | <i>wangkanytja</i> | <i>patjantja</i> | <i>pungkunyitja</i> | <i>tjunkunyitja</i> |

imperfective, future, imperative, imperative imperfective, and characteristic. In addition, there are serial and nominalized verb forms. Each verbal category is manifested by up to four different allomorphs (e.g., imperative: -Ø, -*la*, -*wa*, -*ra*), depending on the conjugational class. The P/Y system is economically analyzed in terms of three stem types: a simple stem which functions as a base for perfective categories, an augmented stem for imperfective categories, and an additional augmented stem for the aspect-neutral forms: see Table 3. The augmented forms were probably inflected words in an earlier stage of the language, with the present-day forms resulting from 'double-marking.'

The Ø-class and l-class are open, with predominantly intransitive and transitive memberships, respectively. The ng-class and n-class are likewise predominantly intransitive and transitive respectively, but they have only a handful of basic roots each. These roots, furthermore, are the only monosyllabic verb roots in the language: n-class: *ya*- 'go,' *tju*- 'put,' *ma*- 'get'; ng-class: *pu*- 'hit,' *nya*- 'see' and *yu*- 'give' (examples from Yankunytjatjara). The overall membership of the ng-class and n-class is very large, however, because numerous verbs are formed by compounding with the basic roots or via derivational affixation. Derivational processes are sensitive to mora parity, as well as to the transitivity preference of the verb class. For example, the main intransitive verbaliser is suffix -*ri/-ari*. The derived stem belongs to the ng-class if it has an even number of morae, and to Ø-class if it has an odd number of morae.

Complex Sentences

A single clause may contain more than one verb, if the subsidiary verbs are suffixed with the serial ending. It is common in narratives for clauses to contain several serial verbs, as well as the main finite verb. The grammar of serial verbs and their associated NPs and modifiers is quite complex. Typically for WDL, subordinate clauses are formed by adding case suffixes to a nominalized clause. For example, a purposive clause is formed with suffix -*ku* (identical with purposive case), e.g., *kungka-ngku mai pau-ntja-ku* [woman-ERG food bake-NOML-PURP] 'so the woman could cook food.' Inside the subordinate clause, the subject, object, and any other NPs occur with the same case-marking as they would have in a simple clause. The circumstantial clause is formed in Yankunytjatjara with suffix -*la* (one of the locative suffixes), e.g., *kungka-ngku mai pau-ntja-la* [woman-ERG food bake-NOML-LOC] 'while/because the woman cooked the food.' The Pitjantjatjara circumstantial is *nya-ngka*, which has likely descended from an earlier *-*nyitja-ngka* (simplification of the first of two nasal-stop clusters is common in WDL phonology). Another subordinate type is the aversive clause, which identifies an outcome to be avoided or prevented.

P/Y purposive and circumstantial clauses comply with a 'switch reference' constraint, i.e., they can only be used if the subordinate clause subject refers to a different individual to the main clause subject. If the subjects are the same, a different subordinate

structure is used in place of the purposive, with the ‘intensive’ suffix *-kitja*. An interesting feature of the intensive construction is that the clause as a whole takes an ergative suffix (*-ngku*) if the verb of the main clause is transitive, e.g., *mai pau-ntji-kitja (-ngku)* [food bake-NOML-INTENT-(ERG)] ‘(wanting) to cook food.’ ‘Actor agreement’ of this kind is also found with adverbs of manner and emotion (better regarded as ‘active adjectives’), and with frequency expressions.

There are three coordinating conjunctions: *ka* ‘and, but,’ *munu* ‘and,’ and *palu* ‘but, even though.’ Unusually for Australian languages, switch-reference operates for coordination. Normally, *ka* can only be used as a conjunction if the subject of the new clause refers to a different individual to the subject of the preceding clause; otherwise, *munu* is used. A range of free and clitic particles express illocutionary and discourse-related meanings.

Pitjantjatjara Text Extract. From *Wati Tjan-garangu Iti Intiritjunanyi* ‘There’s an Ogre Pinching the Baby!’, told by Anmanari Alice. Revised edition published by NW Resource Centre, Ernabella.

Ka-l minyma-ngku panya pata-ra
CONTR-QUOT woman-ERG THAT.ONE wait-SERIAL
watja-nu.
tell-PAST
‘Then the waiting woman told him.’

“*Panya tjangara-na pungku-la wanti-kati-ngu.*”
that.one ogre-1SG:ERG hit-SERIAL leave-PROCESS-PAST
“‘I killed that ogre and got away.’”

Munu “Nyangatja-na puli-ngka nyina-nyi,
ADD this-1SG:NOM hill-LOC sit-PRES,
nyuntu-mpa pata-ra.”
2SG:NOM-PURP wait-SERIAL
“‘I’ve been sitting here on the hill, -waiting for you
(to get back).’”

Ka wangka-ngu, “Palya nyangatja-n
CONTR say-PAST good here-2SG:ERG
pu-ngu. Munu-li-nku a-ra-lta.”

hit-PAST ADD-IDU-REFL go-IMP-and.then
‘He replied, “You did well to kill it here. Let’s get
out of here.”’

Munu pula ma-pitja-ngu ngura
ADD 3DU:NOM away-go-PAST place
kutjupa-kutu.
other-ALL
‘And so away they went to some other place.’

Bibliography

- Bowe H (1990). *Categories, constituents, and constituent order in Pitjantjatjara, an Aboriginal language of Australia*. London: Routledge.
- Douglas W H (1958/1964). *An introduction to the Western Desert Language*. Sydney: Oceania Linguistic Monographs No 4 [Revised].
- Eckert P & Hudson J (1988). *Wangka wiru: a handbook for the Pitjantjatjara language learner*. Underdale: South Australian College of Advanced Education.
- Goddard C (1986). *Yankunytjatjara grammar*. Alice Springs: Institute for Aboriginal Development.
- Goddard C (1990). Emergent genres of reportage and advocacy in the Pitjantjatjara print media. *Australian Aboriginal Studies* 2, 27–47.
- Goddard C (1992). Traditional Yankunytjatjara ways of speaking – a semantic perspective. *Australian Journal of Linguistics* 12(1), 93–122.
- Goddard C (1996). *Pitjantjatjara/Yankunytjatjara to English dictionary*. [Revised 2nd edn.]. Alice Springs: Institute for Aboriginal Development.
- Goddard C & Kalotas A (eds.) (1985). *Punu: Yankunytjatjara plant use*. Sydney: Angus and Robertson. [Reprinted, IAD Press, Alice Springs, 1995, 2002.].
- Klapproth D M (2004). *Narrative as social practice: Anglo-Western and Australian Aboriginal oral traditions*. Berlin: Mouton de Gruyter.
- Langlois A (2004). *Alive and Kicking: Areyonga Teenage Pitjantjatjara*. Canberra: Pacific Linguistics.
- Rose D (2001). *The Western Desert code: an Australian cryptogrammar*. Canberra: Pacific Linguistics.

Polish

R A Rothstein, University of Massachusetts,
Amherst, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Polish belongs to the Lechitic subgroup of the West Slavic languages, together with the extinct Polabian language and Kashubian, which is often treated as a dialect of Polish (see section on dialectology). It is the native language of most of the nearly thirty-nine

million residents of Poland and of a few million additional speakers living outside of Poland (primarily in the neighboring countries, but also in North America, Australia, and other areas).

Orthography

Like other Slavic languages that were historically in the cultural sphere of the Western Church, Polish uses the Latin alphabet. It did not, however, adopt the

Hussite spelling reforms of the 15th century. Instead, it uses a combination of digraphs and diacritic marks in a system devised by 16th-century printers in Cracow and based in part on pre-Hussite Czech orthography. Thus, voiced and voiceless alveolar fricatives and affricates are represented, respectively, by *ž* (or *rž* when derived from an etymological *r*), *sz*, *dž*, and *cz*. The letter *ł*, which once indicated a dental lateral, now represents a labio-velar glide. For most speakers, there is no distinction between *ch* and *h*, which both represent the voiceless velar fricative; the letter *h* once indicated a voiced velar fricative. Voiced and voiceless palatal fricatives and affricates are represented, respectively, by *zi*, *si*, *dzi*, and *ci* when followed by a vowel and by *ź*, *ś*, *dź*, and *ć* otherwise. The palatal nasal has a similar double representation: *ń/n*. Palatalized labials (or, for some speakers, labial plus palatal glide), which occur only before vowels, are represented by *bi*, *pi*, *mi*, *wi*, and *fi*; the combinations *ki*, *gi*, and *chi* stand for fronted variants of the corresponding velars. The letter *ó* represents a high rounded back vowel derived from an etymological *o*, while the letters *ę* and *a*, respectively, represent front and back mid-nasal vowels or their positional variants (see next section).

Phonology

The Polish phonemic inventory consists of 33 consonantal segments and seven vocalic segments. In addition to the two nasal vowels mentioned above, there are five oral vowels, the basic phonetic realizations of which are [i], [ɛ], [ɐ], [ɔ], and [u] (orthographic *i*, *e*, *a*, *o*, and *u/ó*). Orthographic *y* ([ɨ]) represents an allophone of /i/. The nasal vowels are diphthongal, consisting of [ɛ] or [ɔ] plus a nasal segment: the homorganic nasal consonant before a stop or affricate and a nasalized glide before a fricative. At the end of a word before a pause, the front nasal vowel loses its nasal segment; both nasal vowels do so before orthographic *l* and *ł*.

The consonants that arose from the historical palatalization of velars or from the deiotation of clusters consisting of dental stop or fricative plus glide have lost their palatal character. The historical palatalization of dental consonants, on the other hand, has given rise to a series of palatal affricates and fricatives. As in most other Slavic languages, final voiced obstruents lose voicing before pause. In obstruent clusters, both within phonological words and between words, there is regressive assimilation with respect to voicing; orthographic *rž* and *w* exceptionally devoice following a voiceless consonant within the same word. Before a word-initial vowel or sonorant, a word-final obstruent is voiced in some areas of

Poland (e.g., Cracow, Poznań) and voiceless in others (e.g., Warsaw). This sandhi rule does not affect the pronunciation of prepositions, but does affect the pronunciation of consonants preceding some verbal clitics.

Word stress is normally on the penultimate syllable and can thus fall on a preposition if the following noun or pronoun is monosyllabic, e.g., *pod nim* 'under it.' Some traditional exceptions to the penultimate principle (e.g., words of Latin or Greek origin such as *gramatyka* 'grammar' with antepenultimate stress or certain verbal forms) are normally regularized in the pronunciation of younger speakers. Unstressed vowels are not reduced. There is a growing tendency, especially in emphatic speech, to shift stress to the initial syllable.

Morphology

Nouns distinguish seven cases (nominative, accusative, genitive, dative, instrumental, locative, and vocative), although the vocative is commonly replaced by the nominative, except in titles (e.g., *panie profesorze*, literally 'Mr Professor'). There are also no special vocative forms in the plural or for personal pronouns, and case syncretism reduces the number of distinct forms. Three genders (masculine, feminine, and neuter) are distinguished in the singular by agreement phenomena, and a masculine animate subgender can also be distinguished by its syncretism of accusative and genitive. Certain classes of semantically inanimate masculine nouns also show the accusative-genitive syncretism (e.g., names of dances, monetary units, and mushrooms: *tańczyć mazurę* 'dance a mazurka'; *zapłacić dolara* 'pay a dollar'; *znaleźć borowika* 'find a boletus mushroom').

In the plural there is a binary distinction of masculine-personal (nouns referring to male human beings) and nonmasculine-personal (all other nouns); they are distinguished by the nominative endings, by agreement phenomena, and by the accusative-genitive syncretism of the former vs. the accusative-nominative syncretism of the latter. Some nouns have only plural forms (e.g., *drzwi* 'door[s]'); others are used primarily in the singular (e.g., mass and abstract nouns) but have potential plural forms, which usually acquire specialized meanings (e.g., *wino* 'wine' vs. *wina* 'kinds or portions of wine'; *miłość* 'love' vs. *miłości* 'love affairs'). Adjectives, third-person pronouns, and the past-tense forms of verbs also distinguish three genders in the singular and two in the plural.

Noun declensions are largely gender-based. The masculine and neuter declensions have most endings in common in the singular, while the two feminine

declensions in the singular (for nouns ending in *-a* in the nominative singular and those ending in a consonant, i.e., with zero-ending) also share most endings. There is a class of masculine nouns ending in *-a* that follow the feminine *a*-declension in the singular; all of them refer to male human beings. In the plural, only the nominative, accusative, and genitive endings are partly gender-based; the other case endings are common for all nouns. Some case forms involve mutation of the final stem consonant, and certain case endings are dependent on the nature of that final stem consonant – whether it is ‘soft’ (palatal or ‘historically soft,’ i.e., the result of historical palatalization or deiotation) or not.

Polish verbs belong to one of two aspectual categories: perfective or imperfective. There are also some biaspectual verbs (e.g., *abdykować* ‘abdicate,’ *ranić* ‘wound’). Perfective verbs express accomplishments or transitions; imperfective verbs express states or activities/processes. Imperfective verbs are typically unprefixed; adding a prefix perfectivizes the verb, while sometimes also adding an additional semantic component (e.g., *pisać* ‘write = engage in the activity of writing’/napisać ‘write = get something written’ vs. *przepisać* ‘rewrite,’ *opisać* ‘describe,’ or *popisać* ‘write a little or for a while’). There are also productive ways of imperfectivizing a perfective verb through a change in suffix and/or the stem (e.g., *przepisywać* ‘engage in the activity of rewriting,’ *opisywać* ‘engage in the activity of describing’). Occasionally, corresponding verbs are based on different stems (e.g., imperfective *brać* vs. perfective *wziąć* ‘take’), and some verbs have no corresponding verb of the opposite aspect (e.g., imperfective *mieć* ‘have’ or perfective *zdołać* ‘manage [to do something]’).

Imperfective verbs have synthetic forms for past and present tense and analytic forms for the future tense; perfective verbs form their past tense in the same way as imperfective verbs, but the forms that look like the present-tense forms of imperfective verbs normally express future tense (or, under certain circumstances, potentiality). Analytic forms expressing a pluperfect tense are rare in the contemporary language. The perfective/imperfective distinction is also present in infinitives, imperatives, and conditional/subjunctive forms. Imperfective verbs form verbal adjectives and adverbs expressing simultaneity, while perfective verbs form only verbal adverbs that express temporal precedence or subordination to the action of the main verb. Both perfective and imperfective transitive verbs form passive participles, which can be used with *być* ‘be’ to form passives of state (e.g., *W 1945-tym roku Warszawa była zniszczona* ‘In 1945 Warsaw was destroyed [was in a state of destruction]’) and with

zostać ‘become’ to form passives of action (*Podczas wojny Warszawa została zniszczona* ‘During the war Warsaw was destroyed [they destroyed Warsaw]’).

Within the imperfective aspect, a further distinction is made between determinate and indeterminate verbs of motion. Determinate verbs designate motion in a single direction on a single occasion, while indeterminate verbs do not have those restrictions and can therefore designate repeated motion, the ability to move, etc. (e.g., determinate *iść* vs. indeterminate *chodzić*). Many imperfective verbs also have derived iteratives that express repeated, often regular, actions (e.g., *grywać* ‘play frequently’ from *grać* ‘play’).

Declension, conjugation, and derivation all may involve consonant and vowel alternations, e.g., *miasto* ‘city’ vs. *w mieście* ‘in the city’; *idę* ‘I am going’ vs. *idziesz* ‘you are going’; *ręka* ‘hand’ vs. *rączka* ‘little hand’ or ‘handle.’

Syntax

Polish word order is relatively free and is used, together with sentence intonation, to express the informational structure of the utterance. Thus, the rheme normally follows the theme in emotionally neutral speech. Pronominal and some verbal clitics traditionally follow the first stressed word in a sentence, but this is less true in current usage, especially of the particle *się*, which is historically the enclitic accusative form of the reflexive and reciprocal pronoun. The reciprocal function is still present (e.g., *znamy się* ‘we know one another’), but true reflexive uses are rare (e.g., *bronić się* ‘defend oneself’). The particle has assumed a variety of functions in association with verbs, and in contemporary speech it often immediately precedes or follows the relevant verb, regardless of its position in the sentence. Verbs with *się* can express, among other things, a kind of middle voice (e.g., *myć się* ‘wash/wash up/get washed’) and also an intransitive verb with an unaccusative subject (e.g., *lekcja się zaczyna* ‘class is beginning’). In colloquial speech there is also an enclitic dative reflexive/reciprocal pronoun (*se*).

As in some other Slavic languages, *się* has acquired the function of a generic human subject, parallel to German *man* or French *on*, with third-person singular agreement; only in Polish is there the possibility of a direct object in the accusative (e.g., *tu się rzadko ogląda telewizję* ‘they/people rarely watch television here’). Polish shares with Ukrainian an active verbal construction (based on a form derived from a past passive participle) used to express an action in the past performed by a definite but unspecified human actor (e.g., *zrobiono pomyłkę* ‘they made a mistake/a mistake was made’).

First- and second-person subject pronouns are normally used only for contrast or emphasis; third-person subject pronouns are typically dropped after their first use, unless a previous theme has been reintroduced. Subject pronouns are used in non-familiar address, where the words for *you* (masculine singular *pan*, feminine singular *pani*, mixed group plural *państwo*, etc.) take third-person agreement.

Lexicon

It has been estimated that some 76% of the Polish vocabulary was either inherited from Proto-Slavic or was created within the Polish language. The earliest foreign borrowings came together with Christianity from the Czech lands and included both religious terminology and other words that reflect Czech phonology, rather than the expected Polish derivatives from Proto-Slavic (e.g., *wesoły* ‘merry’ instead of the expected *wiesioły*). Over the course of centuries, however, the major donor language was Latin, followed by French, Greek, and German. Italian and Ukrainian also contributed, as did Russian and English; the influence of the last two became especially strong following World War II. Currently most neologisms come from English or from Latin- or Greek-based internationalisms: a recent example of a semantic calque from English is the use of the words *niedźwiedź* ‘bear’ and *byk* ‘bull’ in the sense of stock-market pessimists and optimists, respectively.

History

The Polish language was first documented in the form of 410 personal and geographical names included in the Latin text of a 12th-century papal bull to the archbishop of Gniezno. The next century brought the first recorded complete sentence, quoted in the text of a Latin chronicle. By the 14th century continuous texts in Polish had been created, and thanks to the efforts of Cracow printers at the beginning of the 16th century, a more or less standardized language appeared. This literary language did not have a clear dialect base, since it included features characteristic of the dialects of the two early political and cultural centers, Gniezno/Poznań (Wielkopolska) in the west and Cracow (Małopolska) in the southeast. It has been suggested that because of the role of Bohemia in the Christianization of Poland, the Czech language served as a point of reference for choosing between features from the two Polish dialect areas.

Polish eventually won the competition with Latin as a literary medium (the 16th-century writer Mikołaj Rej proclaimed to the world that “Poles do not gaggle

like geese – they have their own language”) and also survived the assimilatory efforts of the partitioning powers of the late 18th century (Prussia and Russia). In the period since World War II, the standard language has acquired a much broader social base as well as a vastly expanded technical and specialized vocabulary.

Dialectology

The major dialect areas correspond to historical-geographic regions of Poland: Małopolska in the southeast, Mazowsze in the northeast, Wielkopolska in the northwest, Silesia in the southwest, and Kaszuby along the Baltic coast (north and west of Gdańsk). Although the Polish spoken in the pre-World War II eastern Polish territories (the so-called *kresy*, now part of Lithuania, Belarus, or Ukraine) was distinctive, it was not considered a separate dialect, but the resettlement of many speakers from that area in the territories in the west and north acquired from Germany after the War led to the creation of what are called ‘new mixed dialects.’ The major dialects are traditionally distinguished on the basis of their consonantism: the presence or absence of distinct dental, alveolar, and palatal consonants, and the treatment of obstruents before word-initial vowels or sonorants. The reflexes of historical long and nasal vowels and various morphological criteria are among the features used to make finer dialect distinctions. The dialects of the Kaszuby region are most different from standard Polish and from other Polish dialects, which has led some (mostly non-Polish) linguists to consider Kashubian a separate language rather than a dialect of Polish. Despite the absence of any apparent Kashubian national identity, there have been attempts to establish a Kashubian literary standard.

Bibliography

- Bański P & Willim E (eds.) (2003). *Journal of Slavic Linguistics* 8. *Special Issue on Polish Linguistics*.
 Bethin C Y (1992). *Polish syllables: the role of prosody in phonology and morphology*. Columbus, OH: Slavica.
 Domanski T E (1988). *Grammaire du polonais: phonologie, morphologie, morphophonologie*. Saint-Denis: Presses Universitaires de Vincennes.
 Dziwirek K (1994). *Polish subjects*. New York: Garland.
 Gladney F Y (1983). *Handbook of Polish*. Urbana, IL: G & G Press.
 Grappin Henri (1993). *Grammaire de la langue polonaise* (3rd edn.). Paris: Institut d'études slaves.
 Grochowski M & Hentschel G (eds.) (1988). *Funktionswörter im Polnischen*. Oldenburg: Bibliotheks- und Informationssystem der Universität Oldenburg.

- Hentschel G & Laskowski R (eds.) (1993). *Studies in Polish morphology and syntax*. Munich: O. Sagner.
- Kannenberg G (1988). *Die Vokalwechsel des Polnischen in Abhängigkeit von Flexion und Derivation: eine generative Beschreibung*. Munich: O. Sagner.
- Kański Z (1986). *Arbitrary reference and reflexivity: a generative study of the Polish pronoun się and its English equivalents*. Katowice: Uniwersytet Śląski.

- Nilsson B (1982). *Personal pronouns in Russian and Polish: a study of their communicative function and placement in the sentence*. Stockholm: Almqvist & Wiksell.
- Rothstein R A (1993). 'Polish.' In Comrie B & Corbett G G (eds.) *The Slavonic languages*. London & New York: Routledge. 686–758.
- Szpyra J (1989). *The phonology–morphology interface: cycles, levels, and words*. London & New York: Routledge.

Pomoan Languages

S McLendon, City University of New York, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

The seven Pomoan languages are or were spoken north of San Francisco, in the many verdant valleys of the Coast Range mountains (See **Figure 1**). Especially densely populated were the large valleys through which the Russian River runs and those around Clear Lake, as were the foothills of the Coast Range in the south around Santa Rosa and Sebastopol.

Names

There was no single 'Pomo' tribe or language, although maps and authors frequently so indicate. Each of the seven languages was spoken by residents of at least one, and usually several, politically independent towns, of which some 75 are known. By 2004, these have become amalgamated into 19 distinct federally recognized tribes. Speakers of the seven languages did not have a single name for themselves or for the family of languages as a whole. The name 'Pomo,' which now has that function, was first used to refer to this family by Stephen Powers (1877: 5, 146), and has become increasingly used in the 20th century. It derives from two distinct but similar sounding Northern Pomo terms, one the name of an earlier single town (See McLendon and Oswalt, 1978 for details).

English names for the individual languages were developed by Samuel A. Barrett (1908), modeled on native systems of referring to neighboring languages. The language spoken around the modern town of Ukiah, in the center of Pomoan territory, Barrett called Central Pomo. To the north was Northern Pomo; to the northeast on the edge of the Sacramento Valley, Northeastern Pomo. To the east, on the western portion of Clear Lake, was Eastern Pomo, and southeast at East Lake and Lower Lake, Southeastern

Pomo. To the south of Central Pomo were Southern Pomo, and Southwestern Pomo. This last has a native name, *k'ahšá.ya*, anglicized as Kashaya, which is now preferred.

Unfortunately Barrett, in the style of the times, referred to these seven languages as dialects, even though they are distinctly different, mutually unintelligible, languages. This has led to all seven languages commonly being thought to be mere variations of a single language. In fact, speakers of one language could not understand speakers of any of the others without a considerable period of learning, and all but one of these languages were each spoken in several dialects.

Internal Relations

Classifications of the interrelationships of these languages have been proposed by Barrett (1908: 100), Alfred Kroeber (1925: 227), Abraham Halpern (1964: 90), and Robert L. Oswalt (1964: 416). Halpern was the first phonetically competent linguist to collect data on all seven languages. He proposed two slightly different classifications based on sound shifts that he identified but never published. Oswalt (1964: 413–427) based his classification on a comparison of the 100-word lexicostatistical basic word list in each of the seven languages. Halpern and Oswalt agree in identifying Eastern Pomo and Southeastern Pomo as the most divergent, and Southern Pomo and Kashaya as closely related. They differ in the position they assign the geographically isolated Northeastern Pomo, and their conception of the relationship between the three languages spoken in wide contiguous bands from the Russian River to the Pacific: Northern Pomo, Central Pomo, and Southern Pomo. (See **Figures 1 and 2**.)

Significant intermarriage between neighboring towns and the tradition of sending children to be raised by grandparents for extended periods resulted in more than one language being spoken in each

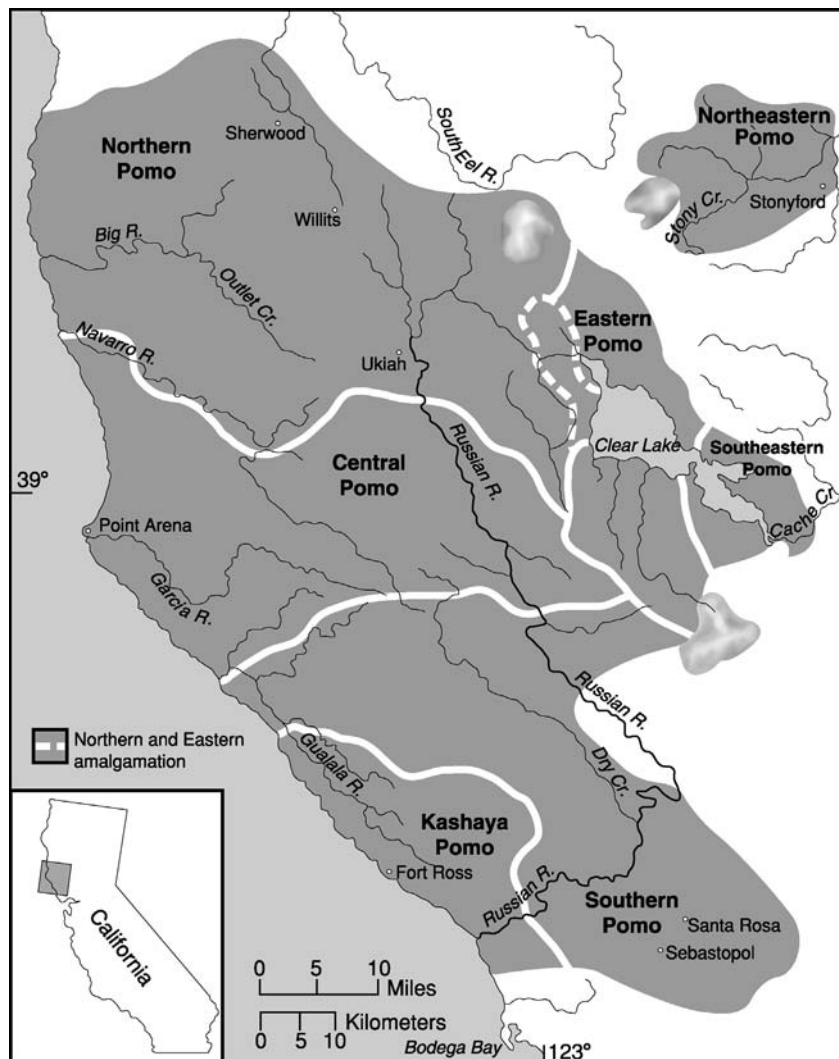


Figure 1 Probable territories of the seven Pomoan languages at the end of the 18th century around the time of first contact with Europeans. Adapted from Figure 2, p. 276 of the *Handbook of North American Indians, California-8* (Washington, D.C.: Smithsonian Institution, 1978).

town, and children having an easy familiarity with more than one language (McLendon, 1978b). This may have had a leveling effect on the languages in contact along the Russian River.

State of Descriptive Knowledge

As of 2004, modern linguistic fieldwork has been carried out on all seven languages, with grammars and articles published on Eastern Pomo (McLendon, 1975, 1978a, 1979, 1982, 1996, 2003), Southeastern Pomo (Moshinsky, 1974), and Northern Pomo (O'Connor, 1984, 1990, 1992). For Kashaya, an extensive unpublished grammar (Oswalt, 1961) exists, as well as articles (Oswalt, 1983, 1986, 1998). Extended field work has been carried out on Central Pomo, with various aspects of the language

described in articles (Mithun, 1988, 1990, 1993, 1998). Extensive fieldwork has been carried out on Southern Pomo by Halpern and Oswalt, very little of which has been published (Oswalt, 1977 provides a text with a grammatical sketch). This is unfortunate, since a clear understanding of Southern Pomo is especially important for the reconstruction of Proto Pomo. The least amount of work, all unpublished, has been done on Northeastern Pomo, which ceased to be spoken in the middle of the 20th century. Ironically, the seven Pomoan languages began to be adequately studied and described only in the mid-20th century, just as speakers were switching to English as their primary means of communication. In 2004, this process of replacement is virtually complete, although several contemporary tribes have now initiated language revitalization efforts.

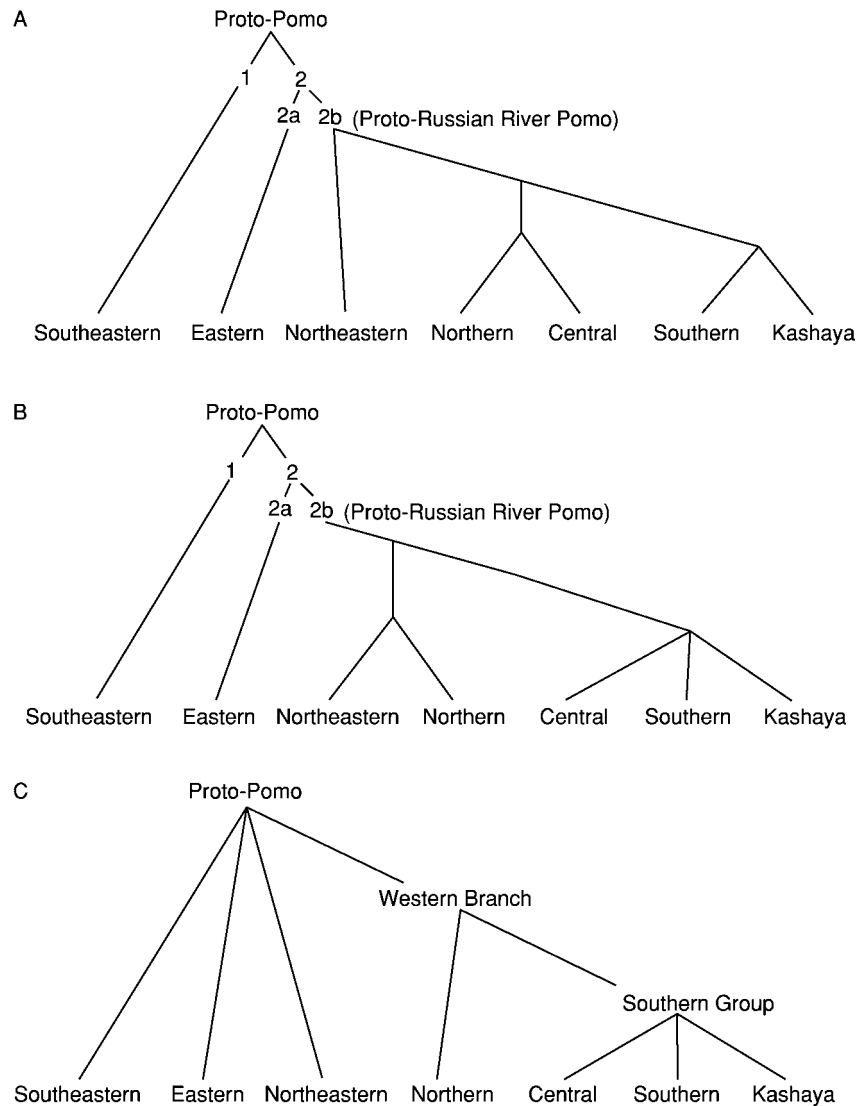


Figure 2 Proposed interrelationships between the seven Pomoan languages. A, B, Two alternative classifications proposed by A.M. Halpern, 1964. C, Classification proposed by R. Oswalt, 1964. After Figure 1, p. 275 of the *Handbook of North American Indians*, California-8 (Washington, D.C.: Smithsonian Institution, 1978).

Basic Characteristics of the Pomoan Languages

Phonology

The seven Pomoan languages have far more consonants than English. Unaspirated, aspirated, and glottalized (or ejective) stops contrast at labial (p , p^b , p^h), dental (t , t^b , t^h), alveolar ($ɬ$, $ɬ^b$, $ɬ^h$), palatal (c , c^b , c^h), velar (k , k^b , k^h), and postvelar (q , q^b , q^h) places of articulation, in Kashaya, Central Pomo and Eastern Pomo (which, however lacks q^b). Compare, for example, the contrasting Eastern Pomo set: *kóy* 'sore,' *k^ból* 'worm,' *kóy* 'in/with the stomach,' *qóy* 'swan,' *q^hoy* 'nape of neck.'

Southeastern Pomo contrasts voiceless stops with glottalized ones at the same places of articulation; voiceless aspirated stops have become fricatives in Southeastern Pomo. Compare Southeastern Pomo *m^hfet*: Eastern Pomo *nur^bér*: Central Pomo *m^bé*: Southern Pomo *nur^bé*: Kashaya Pomo *nup^bé*: Northeastern Pomo *fé*[-*ka*] 'skunk'. Southern Pomo, Northern Pomo and Northeastern Pomo have the same contrasts at the same places of articulation (except Northeastern Pomo has *f* instead of *p^b*), but lack the velar/post-velar distinction (Northern Pomo has no *ɕ^b*). Southern Pomo, Northern Pomo and Eastern Pomo have an additional pre-palatal affricate series (c , c^b , c^h) that pattern like stops (Northern Pomo lacks

c^h). Kashaya and Central Pomo have only *c'*. All seven languages have a glottal stop.

All seven languages distinguish two voiced stops: *b* and alveolar *d*, and the fricatives: *s*, *ʃ*, and *h*. Northeastern Pomo and Southeastern Pomo add a rare *f*, from earlier Proto Pomo **p^h*. Eastern Pomo and Southeastern Pomo have a velar fricative *x*, Southeastern Pomo adds a postvelar fricative *χ*. The sonorants/resonants are *m*, *n*, *l*, *y*, *w* with a rare *r* in Eastern Pomo. Eastern Pomo alone contrasts voiced *m*, *n*, *l*, *y*, *w* with voiceless *M*, *N*, *L*, *Y*, *W*. Compare, for example, *lal* 'month': *Lal* 'goose.' All languages have a five-vowel system with two degrees of length.

Grammar

The seven Pomoan languages are agglutinative, with extensive, complex morphologies and striking semantic specialization. The basic morphological unit is the stem, with verbal or nonverbal function specified by inflectional suffixes and/or syntactic relations. The verb is morphologically the most complex and syntactically the most important category, being the only obligatory member of an independent clause. Verbs are composed of a stem plus a varying number of classes of suffixes that add both lexical and grammatical meaning. These suffixes specify aspects, modes, plurality, locality, reciprocity, source of information (evidentials) and various types of syntactic relations, including the continuation or change in the referent and case roles of the agent and patient of a preceding clause (often called switch-reference). The verb is last in its clause, although under certain conditions, arguments can be postposed following it.

Kashaya and Eastern Pomo have well-developed sets of what have been called instrumental prefixes with the shape CV, where V is *i*, *a*, or *u*. They indicate the undergoer/patient of the action, and type or manner of action, as well as the instrument. These combine with roots to form stems. In Northern Pomo, Central Pomo, Southern Pomo, and Southeastern Pomo, vowels in initial syllables are elided or assimilated, collapsing what are historically several prefixes into a single consonant, obscuring the system.

The Pomoan languages are stative-active languages, most of an unusual type that can be called fluid stative-active. That is, verbs can appear with an argument in either the agentive or the patient case, depending on the speaker's perception of the degree of control the protagonist had in the action. Thus in Eastern Pomo, one can say:

ha c'exél-k-a
1SG.AG slip/slide-PUNCTUAL-DIRECT
'I'm sliding (as on a sled or skis,
 deliberately)'

or:

wi c'exél-k-a
1SG.Pat slip/slide-PUNCTUAL-DIRECT
'I'm slipping (accidentally, as on a
 banana peel, or patch of ice)'

Clauses are combined to describe interrelated sequences of events by affixing one of a number of so-called switch-reference suffixes that indicate simultaneity, sequentiality, causality, or contingency as well as the continuation or change of the protagonists involved and their case roles. Clauses are nominalized by affixing inflectional case marking at their end.

All seven languages identify the sorts of evidence on which an assertion is based, but they differ in the number of distinctions made and the forms used to make them. In Eastern Pomo, for example, suffixes distinguish claims based on (a) direct sensory evidence, (b) someone else's reporting, (c) inferences from circumstantial evidence, or (d) direct knowledge. Evidentials were especially elaborated in Kashaya, Southern Pomo, and Central Pomo (see McLendon, 2003 for details).

Among nonverb classes, kinship terms and pronouns always refer to human animates and are inflected for several cases: agent, patient, possessive, usually commitative, and in some languages, vocative. Personal names existed, but were not used in address or polite reference, an appropriate kin term being preferred. When kin terms were not appropriate, a small closed class of nouns referring to humans of both sexes in various age grades (boy, girl, young lady, young man, man, woman, old man, old woman), usually having suppletive plural stems, were used.

Historical Relationships

Many cognates can be found between the seven languages, demonstrating clear sound correspondences. These usually involve small shifts in sound: either adjustments in place of articulation – postvelars becoming velars, for example, or in manner of articulation – aspirated voiceless stops becoming fricatives. Much more sweeping in their effects are the prosodically conditioned syntagmatic changes that largely affect vowels in particular positions.

If one only looks at lexical comparisons, the languages seem extremely close. However, they show

considerable differences in grammatical structure. When the same category exists, it is frequently expressed by a totally different, not cognate, form in the various languages. When languages have reflexes of the same morpheme, that morpheme may well behave in quite different ways or occur in different relative positions (see McLendon, 1973 and Oswalt, 1976 for details).

Bibliography

- Barrett S A (1908). *The ethnogeography of the Pomo and neighboring Indians*. In University of California Publications in American Archaeology and Ethnology, 6(1). Berkeley: University of California Press. 1–332.
- Halpern A M (1964). A report on a survey of Pomo languages. In Bright W (ed.) *Studies in Californian linguistics*. Los Angeles: University of California Press. 88–93.
- Kroeber A L (1925). Handbook of the Indians of California. *Bureau of American Ethnology, Bulletin* 78, 1–995. Washington, DC: Smithsonian Institution.
- McLendon S (1973). *Proto Pomo*. In University of California Publications in Linguistics, 71. Berkeley: University of California Press.
- McLendon S (1975). 'A grammar of Eastern Pomo.' In University of California Publications in Linguistics, 75. Berkeley: University of California Press.
- McLendon S (1978a). 'Ergativity, case and transitivity in Eastern Pomo.' *International Journal of American Linguistics* 44, 1–9.
- McLendon S (1978b). 'How languages die: a social history of unstable bilingualism among the Eastern Pomo.' In Langdon M, Klar K & Silver S (eds.) *American Indian and Indo-European studies: papers in honor of Madison Beeler*. The Hague: Mouton. 137–150.
- McLendon S (1979). 'Clitics, clauses, closures and discourse in Eastern Pomo.' In *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistic Society*. Berkeley: Berkeley Linguistic Society. 637–646.
- McLendon S (1982). 'Meaning, rhetorical structure and discourse organization in myth.' In Tannen D (ed.) *Georgetown University Round Table on Languages and Linguistics, 1981: analyzing discourse. Text and talk*. Washington, DC: Georgetown University Press. 284–305.
- McLendon S (1996). 'A sketch of the Eastern Pomo language.' In Sturtevant W C (gen. ed.) *Handbook of North American Indians, vol. 15: Languages*. Goddard I (vol. ed.). Washington, DC: Smithsonian Institution. 274–288.
- McLendon S (2003). 'Evidentials in Eastern Pomo with a comparative survey of the category in other Pomoan languages.' In Aikhenvald A & Dixon R M W (eds.) *Studies in evidentiality*. Amsterdam/Philadelphia: John Benjamins. 101–129.
- McLendon S & Oswalt R L (1978). 'Pomo introduction and synonymy.' In Sturtevant W C (gen. ed.) *Handbook of North American Indians, vol. 8: California*. Heizer R (vol. ed.). Washington, DC: Smithsonian Institution. 507–550.
- Mithun M (1988). 'Lexical categories and number in Central Pomo.' In Shipley W (ed.) *In honor of Mary Haas*. Berlin: Mouton. 517–537.
- Mithun M (1990). 'Third-person reference and the function of pronouns in Central Pomo natural speech.' *International Journal of American Linguistics* 56, 361–376.
- Mithun M (1993). 'Switch-reference: clause combining in Central Pomo.' *International Journal of American Linguistics* 59, 119–136.
- Mithun M (1998). 'Fluid aspects of negation in Central Pomo.' In Hinton L & Munro P (eds.) *American Indian languages: description and theory*. Berkeley: University of California.
- Moshinsky J (1974). *A grammar of Southeastern Pomo*. University of California Publications in Linguistics 72.
- O'Connor M C (1984). 'Case marking and third person point of view in Northern Pomo.' *Proceedings of the Pacific Linguistic Conference* 1, 225–246.
- O'Connor M C (1990). 'Third person reference in Northern Pomo conversation: the indexing of discourse genre and social relations.' *International Journal of American Linguistics* 56, 377–409.
- O'Connor M C (1992). *Topics in Northern Pomo grammar*. New York: Garland.
- Oswalt R (1961). *A Kashaya grammar*. Ph.D. diss., University of California, Berkeley.
- Oswalt R (1964). 'The internal relationships of the Pomo-family of languages.' *Actas y Memorias de XXXV Congreso Internacional de Americanistas* 2, 413–427.
- Oswalt R (1976). 'Comparative verb morphology of Pomo.' In Langdon M & Silver S (eds.) *Hokan Studies*. The Hague: Mouton. 13–28.
- Oswalt R (1977). 'Retribution for mate-stealing [Southern Pomo text with grammatical sketch].' In Golla V & Silver S (eds.) *Northern California texts. International Journal of American Linguistics, Native American Texts Series* 2. 71–81.
- Oswalt R (1983). 'Interclausal reference in Kashaya.' In Haiman J & Munro P (eds.) *Switch-reference and universal grammar*. Amsterdam/Philadelphia: John Benjamins. 267–290.
- Oswalt R (1986). 'The evidential system of Kashaya.' In Chafe W & Nichols J (eds.) *Evidentiality: the linguistic coding of epistemology*. Norwood, NJ: Ablex. 29–45.
- Oswalt R (1998). 'Three laryngeal increments of Kashaya.' In Hinton L & Munro P (eds.) *American Indian languages: description and theory*. Berkeley: University of California. 87–94.
- Powers S (1877). *The tribes of California. Contribution to North American Ethnology* 3. Washington, DC: U.S. Geographical and Geological Survey of the Rocky Mountain Region.

Portuguese

A T de Castilho

© 2006 Elsevier Ltd. All rights reserved.

Portuguese is the fifth most widely spoken language in the world, being spoken in Europe (Portugal), South America (Brazil), and Africa (Angola, Mozambique, São Tomé and Príncipe Islands, Cape Verde, and Guinea-Bissau). Approximately 168 million people speak the language, most of them in Brazil. In Portugal and Brazil, Portuguese is the native language, whereas in the other countries it is the official state language, being native for less than 20% of the population.

History

Portuguese is a Romance language, belonging, with Spanish and Catalan–Valencian–Balear (Catalan), to the Ibero–Romance subgroup (*see Catalan; Romance Languages; Spanish*).

It arose from Vulgar Latin, which was brought to the Iberian Peninsula between 218 and 19 B.C. Once the conquest of the peninsula was an established fact, the Romans divided the new province into two parts: Hispania Ulterior ('Farther Spain,' including Baetica and Lusitania), where Galician (*see Galician*) developed, and Hispania Citerior ('Nearer Spain', including Tarraconensis and Gallaecia), where various linguistic varieties, including Spanish and Catalan, developed. The two regions underwent different forms of colonization. Hispania Ulterior was colonized by the senators of the Roman aristocracy, giving rise to a conservative form of Latin. Hispania Citerior, on the other hand, was colonized by military men, leading to the development of an innovative linguistic variety. This explains in part the differences between Portuguese and Spanish.

The original Latin base was modified by contact with the Germanic tribes who dominated the peninsula from the 5th to the 7th centuries, and with the Arabic tribes who dominated two-thirds of the peninsula from the 8th to the 15th centuries. After an inevitable bilingual phase, Latin emerged victorious, being transformed into a peninsular Romance language after the 8th century.

Portuguese arose in the northwest of the Iberian peninsula, specifically in the County of Portucale, one of the divisions of the Kingdom of Castile. Initially, Portuguese formed a single language with Galician, although this unity was threatened with the movement of Portuguese to the south during the Reconquest.

The first texts in Portuguese can be divided into literary and nonliterary texts. The earliest nonliterary

texts date from the 13th century. During the reign of D. Dinis (1279–1325), Portuguese became the official language of Portugal and was used to write legal documents. The oldest nonliterary text dates from 1214. It is the *Testamento de D. Afonso II*, the third king of Portugal. The next was *Notícia de Torto*, written between 1214 and 1216, which tells of a disagreement ('*torto*') motivated by the mismanagement of rural property.

The oldest literary texts date from the 12th century: the *Cantiga d'Escárnio* written in 1196 by Joan Soárez de Pávia, the *Cantiga da Ribeirinha* by D. Sancho I, and the *Cantiga de Garvaia* by Pai Soares de Taveirós. Medieval Galician poetry consists of 1679 lyric and satiric poems and 427 religious compositions, written between 1196 and 1350. The prose texts consist of versions of Latin and French literature in translation, historiography, and religious and philosophical texts.

During the commercial expansion in the 15th and 16th centuries, Portuguese was taken to Africa, Asia, and America. In these regions, pidgins arose and some of these became creoles.

Portuguese pidgins were the first Romance pidgins to emerge. They developed principally in western Africa from the last quarter of the 15th century in Cape Verde, Sierra Leone, the islands of São Tomé and Príncipe, and Guinea-Bissau. Curiously enough, these pidgins were developed in Europe itself during the training of Africans brought to Portugal to learn the language so that they could act as interpreters for the merchants.

These pidgins gave rise to creoles throughout the world. In Africa, there are various creoles, including those of São Tomé and Príncipe (Angolar, Forro, Principense (Moncó), Cape Verde, and Guinea-Bissau. In Asia, the semicreole Sino-Portuguese of Macao was further influenced by Portuguese, whereas the Malayan Portuguese of Java, Malacca, and Singapore, and the Indian Portuguese of Sri Lanka, Goa, Damao, and Diu have almost disappeared. In the Caribbean, Papiamentu from the island of Curaçao was relexified, and, in the late 20th century, is a creole of Spanish. And in South America from the 17th century, a group of Jews left Brazil with their slaves, taking their creole with them to Surinam (Dutch Guyana).

Characteristics of Portuguese

In both Europe and Latin America, Portuguese-speaking countries are bordered by Spanish-speaking ones; there are, however, a few differences separating the two languages. The following sentences can be used to exemplify some of these differences, as well as

those between European Portuguese (EP) and Brazilian Portuguese (BP).

Portuguese: *A mulher comprou os ovos mais lindos da feira.* (1)

The woman bought the eggs most beautiful of the market.

Se tivesse mais dinheiro, levaria também para sua irmã. (2)

If (she) had more money, (she) would take (some) also to her sister.

Syntactic Characteristics

Not only EP but BP has a preferred SVO word order, as does French. Spanish, however, tends to prefer an OVS order: *Los huevos más lindos de la feria los ha comprado la mujer.*

The subject is omitted in EP (*se Ø tivesse mais dinheiro . . .*) and in Spanish (*si Ø tuviera más dinero . . .*). In BP, however, there is a tendency to repeat the subject: *se a mulher/se ela tivesse mais dinheiro . . .*

The direct object is expressed by an NP or a clitic in EP (*A mulher comprou os ovos/A mulher comprou-os . . .*) and in Spanish (*La mujer ha comprado los huevos/La mujer los ha comprado . . .*), whereas the tonic pronoun may either be used or omitted in BP (*A mulher comprou eles/A mulher comprou Ø*).

Morphological Characteristics

The Verb Portuguese maintains the distinction between the *preterito perfeito simples* ‘simple preterite’ (*comprou*), used to express the perfective aspect, and the *preterito perfeito composto* ‘compound preterite’ (*tem comprado*), used for the imperfect aspect; the auxiliary for the compound tense in Portuguese is *ter*. There is a tendency, however, for the corresponding Spanish forms (*compró* and *ha comprado*) to have lost this distinction; moreover, the auxiliary for Spanish is *haber*. Portuguese distinguishes the *imperfecto do subjuntivo* ‘imperfect subjunctive’ (*tivesse*), which is a subordinate tense, from the *mais que perfeito do indicativo* ‘pluperfect indicative’ (*tivera*), which indicates the distant past. Spanish has lost the *imperfecto do subjuntivo*, replacing it with the *mais que perfeito do indicativo* (*si tuviera más plata*).

The Adjective The comparative degree is formed with reflexes of Latin *magis* in both Portuguese and Spanish, respectively *mais lindos* and *más lindos*, in contrast to the French and Italian reflexes of *plus*, respectively *plus beaux* and *più belli*.

Phonological Characteristics

Monophthongs Portuguese has seven stressed vowel phonemes: /a/, /ɛ/, /e/, /i/, /ɔ/, /o/, /u/. This contrasts

with the five of Spanish, since in Portuguese the half-closed and half-open front and back vowels are used distinctively, as for example in the singular and plural of ‘egg’ (*ovo* /'ovu/, *ovos* /'ɔvus/) and in the masculine and feminine third-person pronouns (*ele* /'ele/, *ela* /'ela/).

Portuguese also developed nasal vowels with phonemic value (*lindo* /'lĩdu/ ‘beautiful,’ *lido* /'lidu/ ‘read’); this did not happen in Spanish.

Diphthongs Spanish diphthongized the short vowels (*övu* > *huevo*), whereas Portuguese did not (*övu* > *ovo*), except in certain dialects. Diphthongs did develop in Portuguese when an intervocalic consonant was eliminated and two vowels within a single word became contiguous; these vowels then occur in Portuguese in words that have simple vowels in Spanish: Portuguese *mais*, Spanish *más*; Portuguese *comprou*, Spanish *compró*; Portuguese *coisa*, Spanish *cosa* ‘thing’; Portuguese *dinheiro*, Spanish *dinero*.

Consonants Portuguese lost intervocalic [n] and [l], whereas Spanish retained them: *irmã/hermana* ‘sister’; *dor/dolor* ‘pain.’

Varieties of Portuguese

EP presents a notable lack of differentiation, with the variety of Lisbon providing the standard. The substitution of [v] for [b], the apico-alveolar pronunciation of [s] and [z], the maintenance of the affricate [tʃ], and the maintenance of the diphthongs [aw] and [ow], distinguish the dialects of the north (Trasmontano, Interamnense, Beirao) from those of the south (Estremenho, Alentejano, Algarvio). In Portuguese territory, various varieties of Leonês are also spoken: Rionorês, Guadramilês, and Mirandês.

The introduction of EP to Brazil began in the 16th century. There it came into contact with the 300 indigenous languages spoken by approximately 1 million individuals, as well as with those of some 18 million Negro slaves from the Bantu and Sudanese cultures who were brought to the country over a period of three centuries. BP went through three historical phases: (a) 1533–1654, a phase of bilingualism with a strong predominance of Tupinambá (Old Tupi); (b) 1654–1880, a phase during which Old Tupi gave way to creole varieties; and (c) after 1808, a phase involving an intense urbanization of the country, with massive immigration of Portuguese settlers and a consequent approximation of BP to EP. This last phase also marked the beginning of the distinction between rural and urban speech.

BP also presents great uniformity, although there are minor differences. The speech of the north (Amazon

and the northeast) is distinguished from that of the south (Mineiro, Paulista, Carioca, and Gaúcho) by the raising of the pretonic medial vowel resulting in the production of a close vowel (*feliz* /fi'liʃ/ 'happy,' *chover* /ʃu'ver/ 'to rain') or by an open vowel (*feliz* /fe'liʃ/, *noturnu* /nɔ'turnu/ 'nocturnal'), by the nasalization of vowels followed by a nasal consonant (*cama* /kâma/ 'bed'), by the replacement of [v] with [b] (*varrer* /ba'rer/, *vassoura* /ba'sora/ 'broom'), and by the affricates /tʃ/ and /dʒ/ (*oito* /'oytʃu/ 'eight,' *muito* /'mūtʃu/ 'too much'). There is no single standard, but rather several centers and regional standards: Belém, Recife, Salvador, São Paulo, Rio de Janeiro, and Porto Alegre. In the south, BP penetrates into Uruguayan territory.

Since the 19th century, the relationship between BP and EP has been an object of attention. Two different hypotheses have been advanced: the creolization hypothesis and the parameter-change hypothesis. According to the first, BP had a pidgin phase, which gave rise to a creole; this is in the early 1990s in the process of decreolization. This hypothesis is strengthened if the written language is taken into consideration, since in schools the attempt is made to make written BP conform closely to written EP. However, an examination of the spoken language makes it impossible to suppose that there has been a change in the direction of EP, which is leading to a syntactic convergence of the two varieties. For this reason, the second hypothesis, parameter change, seems more probable. According to this, BP grammar has diverged from the grammar of EP in the following ways: (a) retention of the subject, which is omitted in EP because it is already reflected in the verbal morphology; (b) progressive loss of subject inversion, maintained in EP; (c) loss of the clitic system of the third person (retained in EP) and object omission; and (d) changes in relativization rules, with the disappearance of the pronouns *cujo* and *onde*, and the appearance of the relative pronoun without a

preposition (*o livro que eu preciso* instead of *o livro de que eu preciso* 'the book I need'), as well as the repetition of the referent of the relative pronoun (*o menino que a casa dele pegou fogo* instead of *o menino cuja casa pegou fogo* 'the boy whose house caught fire'; *a casa que eu nasci lá* instead of *a casa onde nasci* 'the house where I was born'). Further studies, especially in the area of syntax, will shed more light on the precise nature of the differences between BP and EP.

Bibliography

- Camara J M (1975). *História e Estrutura da Língua Portuguesa*. Rio de Janeiro: Padrao.
- Castilho A T de (ed.) (1991). *Gramática do Português Falado 1*. Campinas: Editora da UNICAMP.
- Cintra L F L (1959). *A Linguagem dos Foros de Castelo Rodrigo*. Lisbon: CEF.
- Cunha C & Cintra L F L (1985). *Nova Gramática da Língua Portuguesa Contemporânea*. Rio de Janeiro: Nova Fronteira.
- Holanda-Ferreira A B (1986). *Novo Dicionário da Língua Portuguesa* (2nd edn.). Rio de Janeiro: Nova Fronteira.
- Ilari R (ed.) (1992). *Gramática do Português Falado (2)*. Campinas: Editora da UNICAMP.
- Maia C (1986). *História do Galego – Português*. Coimbra: INIC.
- Mattos e Silva R V (1989). *Estruturas Trecentistas*. Lisbon: Casa da Moeda.
- Mira Mateus M H, Brito A M, Duarte I & Faria I H (1989). *Gramática da Língua Portuguesa* (2nd edn.). Lisbon: Caminho.
- Pimpao A J C (1947). *História da Literatura Portuguesa (1)*. Lisbon: Quadrante.
- Roberts I & Kato M (eds.) (1993). *Português Brasileiro: Uma*. Campinas: Editora da UNICAMP.
- Silva Neto S (1952). *História da Língua Portuguesa*. Rio de Janeiro: Livros de Portugal.
- Spina S (1991). *A Lírica Tavadoresca*. São Paulo: EDUSP.
- Tatalló F (ed.) (1989). *Fotografias Sociolinguísticas*. Campinas: Pontes.

Punjabi

T K Bhatia, Syracuse University, Syracuse, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Punjabi is a modern Indo-Aryan language spoken in primarily two South Asian countries: India and Pakistan and also in countries outside South Asia

(the United Kingdom, Canada, Malaysia, Indonesia, Singapore, Fiji, United Arab Emirates, Kenya, South Africa, and other countries). The name Punjabi (also spelled Panjabi) is derived from Punjab, the land of five rivers (the Jehlam, the Ravi, the Chanab, the Vyas, and the Satluj).

Approximately 50 million people speak Punjabi as either a first or second language. It is the official language of the state of Punjab in India. Although

the official language of Pakistan is Urdu, it is spoken as a native language by just 8% of the population; the majority native language is Punjabi, spoken by approximately 60% of the population. Punjabi is ranked among the top 20 most widely spoken languages in the world.

With the partition of the Indian subcontinent came the partition of the State of Punjab. So massive was the migration in 1947 that it is viewed as the greatest migration in the history of humanity. About 10 million people were uprooted from both sides of what is now India and Pakistan. Consequently, the population of the Punjabi-speaking area underwent radical reorganization, which had and continues to have an impact on the language in term of norms, multiple identities, and language standardization.

History and Literature

Punjabi, which is a descendant of the Sanskrit language, belongs to the Indo-Aryan language family. It has been in use as a literary language since the 11th century. Punjabi has three distinct historical stages: Old (10th–16th centuries), Medieval (16th–19th centuries), and Modern (19th-century to the present). The most important treatise of the Old Punjabi is *A:di grantha*, the sacred scripture of the Sikhs.

Varieties and Dialects

In addition to the national varieties, Punjabi has several regional, religious (Hindu, Sikh, and Muslim), and socioethnic varieties. The four regional varieties of Eastern Punjabi are as follows:

1. Majhi, the standard variety, is spoken in the districts of Amritsar and Gurdaspur.
2. Malwi is found in the districts of Bhatinda, Ferozpur, Ludhiana, the western parts of Patiala and Sangrur.
3. Doabi is spoken in the districts of Jallandar, Kapurthala, and Hoshiarpur.
4. Powadi is dominant in the district of Ropar, and the eastern parts of Patiala and Sangrur.

There are four additional traditionally recognized dialects of Punjabi (Rathi, Ludhianwi, Patialwi, and Bhattani), whose status as independent dialects is subject to dispute. Lahanda (also called Saraiki and Multani), which is classified as Western Punjabi by Grierson (1916), is questioned by language authorities. The Saraiki, Hindoko, and Pothohari (also called Patohari, Pothwari, Putohari, Pothohari, Mirpuri Punjabi) language movements in Pakistan assert the

three varieties as three separate languages in their own right rather than as three dialects of Punjabi (see Rahman, 1996 for details).

Writing Systems

Punjabi is written primarily in three scripts: Gurmukhi, Perso-Arabic, and Devanagari. Sikhs often write Punjabi in Gurmukhi, Hindus in Devanagari, and Muslims in Perso-Arabic, called Shahmukhi. Punjabi written in Gurmukhi is the official language/script of the Indian state of Punjab. In addition to these three scripts, Punjabi is also recognized for its business scripts, such as LaNDa, Mahajani, and Baniyakar. These scripts, although now dying, are particularly noteworthy not only for their telegraphic and ‘shorthand’ characteristics employed in clerical and business domains, but also as a secret code.

Phonology

Punjabi has four notable phonological features.

First, Punjabi is the only modern Indo-Aryan language that has developed tonal contrasts.

- The low tone / ˊ / is characterized as low-rising tone.
- The high tone / ˋ / is characterized as a rising-falling tone.
- The mid tone / - / is never represented by the accent mark since it is predicted by rules of redundancy.

The following examples reflect the phonemic status of the level tones:

/kàR/ ‘chisel’
/kaR/ ‘bottom’
/káR/ ‘boil.’

Although tones are phonemic, none of the three scripts (Gurmukhi, Devanagari, and Perso-Arabic [Shahmukhi]) have any symbol or accent mark to identify tones; instead voiced aspirated consonant symbols are used, which reflect either the older (Eastern) Punjabi or modern Western Punjabi pronunciation. Thus, there is a close correlation of voiced aspirates (e.g., of languages such as Hindi) and the Punjabi tones.

Second, Western Punjabi still retains the original Indo-European (1500 B.C.) distinction between aspirated and unaspirated consonants, which results in a four-way contrast, as shown in the following examples:

ka:l ‘time’
kha:l ‘skin’
ga:l ‘cheek’
gha:l ‘to put into’

Table 1 Consonants

| | | <i>Labial</i> | <i>Dental</i> | <i>Retroflex</i> | <i>Palatal</i> | <i>Velar</i> | <i>Back velar</i> |
|------------|---------------------|---------------|---------------|------------------|----------------|--------------|-------------------|
| Stop | Unvoiced unaspirate | p | t | T | c | k | (q) |
| | Unvoiced aspirate | ph | th | Th | ch | kh | |
| | Voiced unaspirate | b | d | D | j | g | |
| | Nasal | m | n | N | ñ | ŋ | |
| Fricative | Unvoiced | (f) | s | sh | | (x) | |
| | Voiced | | (z) | | | (G) | |
| Flap | Voiced unaspirate | | r | R | | | |
| | Voiced aspirate | | | Rh | | | |
| Lateral | | | l | L | | | |
| Semivowels | | | w(v) | | y | | |

In Eastern/Standard Punjabi, this four-way contrast is reduced to a three-way contrast: unvoiced unaspirate, unvoiced aspirate, and voiced unaspirate. The voiced aspirates yield tones.

Third, it has the feature of retroflexion in its consonant inventory,

Ta:l ‘to put off’

ta:l ‘pond’

The retroflex consonant is transcribed as the capital T. In addition to the retroflex stop, Punjabi has a fricative, flaps, and a lateral.

Fourth, geminates are another distinctive feature of Punjabi. All consonants except N, L, R, y, h, and w may be geminate (doubled). Geminataion is represented by the *addak/adhak* sign in Gurmukhi.

The inventory of distinctive segments of standard Punjabi is as follows. The symbol (·e) indicates the sounds that occur in Perso-Arabic words (Table 1).

Glottal h appears only in the word-initial position.

Punjabi vowels may be oral or nasal. The distinction is phonemic: *ga:* ‘sing’ and *gā:* ‘cow’ (Table 2).

Stress

Although stress (meaning loudness) is not a prominent feature of Punjabi, it seems that its existence cannot be denied. Stress can distinguish between grammatical categories such as nouns and verbs, as in:

| Nouns | Verbs |
|--------------|-----------------------|
| galaa ‘neck’ | galaa ‘cause to melt’ |
| talaa ‘sole’ | talaa ‘cause to fry’ |

The stressed syllable is shown in bold. However, stress is not usually distinctive in Punjabi. Therefore, in general, whether one places stress on the first syllable or on the second, the meaning will not be affected. For example, the meaning of the word *suNa:* ‘heard’ will remain unchanged whether one places stress on the first syllable or the second. Therefore, Punjabi is

Table 2 Vowels

| | <i>Front</i> | <i>Central</i> | <i>Back</i> |
|------|--------------|----------------|-------------|
| High | i: | | u: |
| | i | | u |
| Mid | e | | o |
| | ai | a [schwa] | au |
| Low | | a: | |

often characterized as a syllable-timed language like French, where the syllables are pronounced in a steady flow, resulting in a ‘machine-gun’ effect.

Morphology

Word formation in Punjabi primarily uses prefixes and suffixes to define inflectional and derivational word classes. Nouns are generally inflected for number, gender, and case. There are two numbers, singular and plural; two genders, masculine and feminine; and three cases, simple, oblique, and vocative. The oblique forms occur when a noun or a noun phrase is followed by a postposition. Nouns are inflected according to their gender and word-final sound, as exemplified by the three paradigms given in Table 3.

Adjectives are primarily of three types:

1. Simple adjective, such as *canga:* ‘good’;
2. Derived adjectives employing various parts of speech such as nouns: *marda:na:* (from *mard* ‘man’) ‘masculine,’ adverbs: *manda:* ‘slow’ (from *mand* ‘low’), and from agentive/adjectival particle *va:la:* e.g., *dilli: va:la:* (from *dilli:* ‘Delhi’) *va:la:* ‘-er’) ‘from Delhi’;
3. Participial adjectives: *caldi:* ‘moving,’ *nasda:* ‘running.’

Adjectives can be used both attributively (immediately placed before nouns) and predicatively (immediately placed before verbs). Simple, Participial, and *va:la:* adjectives are of two types, inflected and uninflected. Inflected adjectives agree with their following

Table 3 Punjabi noun paradigms

| Case | Paradigm I: masculine nouns ending in -a: (e.g., muNDa: 'boy') | | Paradigm II: masculine nouns not ending in -a: (e.g., a:dmi: 'man') | | Case | Paradigm III: all feminine nouns (e.g., kuRi: 'girl') | |
|----------|--|---------|---|--------|----------|---|---------|
| | Singular | Plural | Singular | Plural | | Singular | Plural |
| Direct | muNDa: | muNDē | a:dmi: | a:dmi: | Direct | kuRi: | kuRiyā: |
| Oblique | muNDē | muNDiā: | a:dmi: | a:dmiā | Oblique | kuRi: | kuRiyō |
| Vocative | muNDē | muNDio | a:dmi: | a:dmio | Vocative | kuRi: | kuRio |

Table 4 Punjabi pronoun paradigms

| | 1st Person | | 2nd Person | | 3rd Person | |
|----------|------------|--------|------------|----------|------------|----------|
| | Singular | Plural | Singular | Plural | Singular | Plural |
| Direct | māi | asī: | tu: | tussi: | ó | ó |
| Oblique | mā | sa: | tāi | tvā: | ó | ónha: |
| Genitive | mera: | sa:Da: | tera: | tuvā:Da: | óda: | ónha:da: |

noun in number, gender, and case; they end in morpheme *-a:* (e.g., *canga:* 'good'), which changes to *e* for masculine plural and masculine oblique (*cange*), *-ii* for feminine singular (*cangi:*) and *-iā:* for feminine plural nouns (*cangiā:*). Uninflected adjectives (not ending in *-a:*) remain unchanged.

Although the case system of pronouns is essentially the same as that of nouns, pronouns have more case forms than nouns. Case relations are essentially carried out by means of postpositions. Personal pronouns are similar to their English equivalents; however, there are no gender distinctions (like *he* and *she* in English) (Table 4).

Adverbs and postpositions are invariant, except for the genitive postposition, which behaves like an inflected adjective. The postpositions mark case relations and adverbial functions.

Verbs

There are three tenses in Punjabi: present, past, and future. The tenses are formed by the suffixation process. Verbs are inflected for number, gender, and person.

ó a:-nd-a: ai.
he come-PRES-MASC. Sing is
 'He comes.'

ó aa-iyaa.
he come-PERF.MASC.Sing.
 'He came.'

ó aa-e-g-aa
he come-3Sing-FUT-MASC. Sing
 'He will come.'

In addition to simple verbs, Punjabi has two categories termed 'conjunct' and 'complex' verbs. The class of conjunct verbs is usually derived adding *karnaa* 'to do' or *hoNaa* 'to be' to noun, adjective, pronoun, or adverb, for example:

kamm 'work,' kamm karNa: 'to work'

canga: 'good,' canga: hoNa: 'to recover'

tez 'fast,' tez karNa: 'to speed up'

Complex verb: likh 'write,' laiNa: 'to take' → likh laiNa: 'to write' (for one's own benefit)

Punjabi is also sensitive to stative/active and volitional/nonvolitional distinction; these four types of distinction are denoted by morphologically related verbs:

khulNa: 'to be opened,' kholNa: 'to open'

TuTNa: 'to be broken,' toRna: 'to break'

Causative verbs are derived by adding *-a:-*, for the simple causative, and *-wa:-* for the double causative, to the stem of the verb.

Compounding is an integral and very productive process of word formation in Punjabi. The noun-noun compounding involves twelve types of compounding. For example, *kha:Na:* 'eating' and *pi:Na:* 'drinking' can be compounded into *kha:N-pa:N* 'life style.'

From the viewpoint of morphological complexity, Punjabi can be classified as an agglutinating language. Derivation of words takes place by the addition of

suffixes to simple or derived stems of major word classes. The process of prefixation is almost exclusively used with nouns and verbs, other word classes rarely participate in this process. The process of suffixation is equally productive with both nouns and verbs.

Syntax

Punjabi is a Subject Object Verb (SOV) language with relatively fixed word order. Interrogative or other sentence types do not introduce any changes in word order. In topicalization and focus structure, however, phrases occurred in marked position, usually initial. It is primarily a head-final language. The verb generally agrees with the subject. In transitive perfective sentences, the third person subject is marked with the *ne* postposition, and the verb agrees with the direct object. As a rule of thumb, the verb never agrees with any constituent, which is marked with a postposition.

Any sentence can be negativized by placing the negative particle *nahii* 'not' in the preverbal position. Punjabi is a Pro-drop language. In the following sentence, the subject can be dropped.

nahī: a:-e-g-a:
not come-3Sing-FUT-MASC. Sing
'(he) will not come.'

Language Contact

Punjabi borrows from Sanskrit, Persian, Arabic, Hindi-Urdu, and recently, English.

Bibliography

- Bahl K C (1969). 'Panjabi.' In Sebeok T (ed.) *Current trends in linguistics*. The Hague: Mouton. 153–200.
- Bhatia T K (2003). 'The Gurmukhi script and other writing systems of Punjab: history, structure and identity.' In Bhaskararao P (ed.) *International Symposium on Indic scripts: past and present*. Tokyo, Research Institute for the Languages and Cultures of Asia and Africa: Tokyo University of Foreign Studies. 181–213.
- Bhatia T K (2000). 'Lexical pronouns and anaphors and pronouns in Punjabi.' In Lust B, Wali K, Gair J & Subbarao K (eds.) *Lexical anaphors and pronouns in selected South Asian languages*. New York: Mouton de Gruyter. 637–714.
- Bhatia T K (1993). *Punjabi: a cognitive-descriptive grammar*. London: Routledge.
- Bhatia T K (1975). 'The evolution of tones in Punjabi.' *Studies in Linguistic Sciences* 5(2), 212–224.
- Gill H S & Gleason H (1969). *A reference grammar of Panjabi*. Patiala: Punjabi University.
- Grierson G A (1916). *Linguistic survey of India*. Vol. 9:1. Delhi: Motilal Banarasidass. (Reprinted 1967.)
- Malik A N (1995). *The phonology and morphology of Panjabi*. Delhi: Munshiram Manoharlal.
- Rahman T (1996). *Language and politics in Pakistan*. Karachi: Oxford University Press.
- Shackle C (1970). 'Punjabi in Lahore.' *Modern Asian Studies* 4(3), 239–267.
- Shackle C (1977). 'Siraiki: a language movement in Pakistan. 4.' *Modern Asian Studies* 11(3), 239–267.
- Singh K (1963). *The history of the Sikhs*. Vol. 1, 1469–1839. Princeton: Princeton University Press.
- Singh K (1966). *The history of the Sikhs*. Vol. 2, 1839–1964. Princeton: Princeton University Press.

This page intentionally left blank



Quechua

W F H Adelaar, Leiden University, Leiden,
The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The name Quechua (also Quichua or Runa Simi) is used for a group of closely related Amerindian languages or dialects, spoken in parts of the Andean states of Argentina, Bolivia, Colombia, Ecuador, and Peru. Although Quechua is traditionally referred to as a language and its local varieties as dialects, substantial local differences often prevent mutual intelligibility. Estimates for the total number of Quechua speakers vary between 7 and 10 million, although the lower number seems to have a firmer statistical basis. In Peru, the number of Quechua speakers ranges between 3 and 4 million; in Bolivia there are about 2 400 000 speakers. Geographically, the Quechua language area does not form a continuum; it is interrupted by large hispanophone and Aymara-speaking regions.

Linguistic reconstruction suggests that, at the end of the first millennium A.D., Quechua was widely spoken along the coast and in the mountains of central Peru with possible extensions into the southern and northern Andean parts of that country. Toward the end of their expansion, at the end of the 15th century, the rulers of the Inca empire adopted as their language of administration a variety of Quechua, referred to by the Spanish conquerors as *La lengua general del Inga* ('The general language of the Inca'). The Incas contributed to the spread of Quechua into outlying areas of their empire (highland Ecuador, northwestern Argentina) by an active policy of forced migrations (*mitimaes*).

The name Quechua came into use in the second half of the 16th century. It was probably derived from the term for temperate altitude zones situated at about 10 000 feet or from the name of a province and ethnic group in the present-day department of Apurímac. The first grammar of Quechua, by Domingo de Santo Tomás, dates from 1560. During most of the colonial period, the Spanish authorities stimulated the use of Quechua as a language of

colonization and evangelization, introducing it to the detriment of many local languages into areas where it had never been used before. A new standard language was created, soon to be replaced as the language of indigenous prestige by the Quechua dialect of the former Inca capital Cuzco. The early 18th century brought a period of Quechua cultural revival and literary activity. It came to an end around 1770, when cultural and linguistic repression by the Spanish rulers initiated the decline of the Quechua language that has continued until recently. In spite of stimulating measures, such as the official recognition of Quechua as a second national language in Peru (1975), the prospects of Quechua have hardly improved.

From a historical and genealogical point of view, Quechua has no proven external relatives. There are phonological, structural, and lexical similarities with the neighboring Aymaran (also called Jaqi or Aru) languages, which indicate a protracted period of interaction between the two groups. Most of these similarities, which also include more than 20% of shared lexicon, do not extend to other languages in the area. The similarities between Quechua and Aymaran have often been interpreted as proof of a common origin (the Quechumaran hypothesis). However, practically all the similarities can be attributed to convergence, making it difficult to distinguish between borrowed and inherited material (see **Aymara** and **Andean Languages**).

Internally, the Quechua family is subdivided into two main groups. The first group (Quechua I or Central Peruvian Quechua) is located in the Andes of central and central northern Peru. One of its most vital dialects is Ancash Quechua, but many other Quechua I dialects are close to extinction. The second group (Quechua II) comprises the dialects spoken in southern Peru (Ayacucho, Cuzco, Puno), some dialects in northern Peru (Cajamarca, Ferreñafe, Chachapoyas, Lamas), and all the Quechua dialects spoken in Bolivia (Apolo, Cochabamba, Potosí, Sucre), Ecuador (Highland and Eastern Lowlands Quichua), Argentina (Santiago del Estero, Jujuy), and Colombia (the Ingano dialect in Caquetá and

Nariño). The largest numbers of Quechua speakers use the Ayacucho, Cuzco, and Puno dialects; Bolivian Quechua; and Ecuadorian Highland Quechua.

Quechua phonology varies according to dialect, but is generally simple. The vowel system consists of three vowels (*a*, *i*, *u*), the high vowels being lowered to (*e*, *o*) next to a uvular consonant. An additional distinction between long and short vowels is found in Quechua I. Stops are generally voiceless. There is a contrast between velar and uvular stops, although the latter have become fricative in many dialects. Some dialects preserve a distinction between palatal and retroflex affricates. Glottalized and aspirated stops are found in the dialects bordering on Aymara (Cuzco, Puno, Bolivian); aspirated stops are also found in Ecuador. In most dialects, stress is predictably located on the penultimate syllable or mora. Quechua word structure does not allow sequences of consonants within a syllable.

Quechua has an agglutinating structure mainly based on suffixation; there are no prefixes at all. The morphology is complex, but regular. (Ecuadorian and Colombian Quechua have a much simplified morphology in relation to the other Quechua dialects.) Words containing as many as eight consecutive suffixes are not exceptional. Verb-final order is obligatory in dependent clauses, and it is the preferred constituent order in full sentences. In noun phrases, all modifiers precede their heads, except for relative clauses in some dialects.

Nouns can be marked for case, number (plural), and person of possessor. The overall structure of the language is nominative-accusative. Objects must be marked for accusative case, unless they precede a nominalized verb. Plural marking on nouns is optional, although not infrequent. In the southern Quechua II dialects, both the number of a nominal referent and of its possessor can be indicated morphologically. The pronominal system and the personal paradigm include a distinction between inclusive and exclusive first-person plural.

Verbs in Quechua exhibit a rich derivational morphology, including causative, applicative, reflexive,

reciprocal, desiderative, and several other options; they are also marked for tense, mood, aspect, speaker orientation ('hither'), personal reference (including the inclusive-exclusive distinction), and number (of subject and/or object). Personal reference not only includes a specification of the subject but also of a direct or indirect object, provided that the latter is a participant in the speech act. In the area of personal reference, tense, and mood, several portmanteau suffixes occur.

Nominalization and direct verbal subordination play a central role in Quechua morphosyntax. Different types of dependent clauses are obtained by combining nominalized verbs with specific case markers. Nominalization is also used to form relative clauses. Subordinated verbs encode a system of switch-reference (i.e., they indicate whether or not the subject of the dependent clause coincides with that of the main clause).

Sentential affixes or enclitics are used to indicate evidentiality (assertion, hearsay, conjecture), topic-comment structure, interrogation and negation, inclusion ('also'), (non-)completion ('already,' 'yet'), emphasis, and several attitudinal functions.

Bibliography

- Adelaar W F H (1977). *Tarma Quechua, grammar, texts, dictionary*. Lisse: The Peter de Ridder Press.
- Adelaar W F H & Muysken P C (2004). *The languages of the Andes*. Cambridge: Cambridge University Press.
- Cerrón-Palomino R (1987). *Lingüística quechua*. Cuzco: Centro Bartolomé de Las Casas.
- Cole P (1982). *Imbabura Quechua*. Amsterdam: North-Holland Publishing Company.
- Mannheim B (1991). *The language of the Inka since the European invasion*. Austin, TX: University of Texas Press.
- Parker G J (1969). *Ayacucho Quechua grammar and dictionary*. The Hague and Paris: Mouton.
- Torero A (1974). *El quechua y la historia social andina*. Lima: Universidad Ricardo Palma.
- Weber D J (1983). *A grammar of Huallaga (Huánuco) Quechua*. Los Angeles: University of California Press.

R

Rhaeto Romance

G Darms, University of Fribourg, Fribourg, Switzerland

© 2006 Elsevier Ltd. All rights reserved.

Since Th. Gartner (1883), the term ‘Rhaeto-Romance’ has been associated with the Romansh dialect of Graubünden, Switzerland; the Ladin dialects of the Dolomitic Alps in South Tyrol, Italy; and the Friulian dialects around Udine, spoken at the northern and eastern border of the Italian speaking area, at the frontier of the German and Slovene areas. Before Gartner’s time, these three dialectal territories had been linked, though under another name (Ascoli, 1873) or with no name yet (Schneller, 1870). The term ‘Rhaeto-Romance,’ however, had already been used before Gartner to describe the Romansh dialects of Graubünden (Diefenbach, 1831) and in 1938 this term even became the constitutional designation for these dialects in Switzerland. The term therefore is not used with consistency.

The language territory of Romansh is part of the canton Graubünden, an administrative subdivision of Switzerland. The first coherent written evidence is a testimony of a witness from a record written in Latin in 1389. The first longer text was written in 1527, but it has survived only in later copies. In the 16th and 17th centuries, different books of exclusively confessional content were published. They were the starting point for four different regional written standards: the two Engadine written standards Puter and Vallader as well as two different Surselvan variants, a catholic one and a protestant one, that were united at the beginning of the 20th century. In the meantime, however, another regional standard had emerged: Surmeiran. To these standards, Sutselvan was added in the first half of the 20th century. This fragmentation, along with the topographic conditions that hardly allowed for direct contact between the different regions, was the reason why the dialects developed separately, so that a certain acclimatization is actually necessary to assure the supra-regional understanding.

In the constitution of Graubünden, Romansh was put on an equal footing with German and Italian,

the two other cantonal languages of Graubünden, already in 1880. It was, however, never extensively used as a language of administration, especially as the lack of a uniform written standard for the Romansh meant that in each case, two regional written standards had to be used for administration. In the schools of the language territory, in which Romansh is usually the only language of instruction until the third form, all five regional written standards are used, though Sutselvan is used in only one school. In 1938, Romansh was granted the status of a national language, but not of an official language. There have also been attempts to create a uniform written standard for the entire Romansh territory, e.g., in 1864 the *Romonsch fusionau*, which failed however, and more recently the *Rumantsch Grischun* (1982), which is accepted on federal as well as cantonal level as a language of administration.

In the 2000 census, the number of speakers who indicated Romansh as main language amounted to only 27 038 in Graubünden and 35 095 in Switzerland; there are 40 168 in the canton (60 651 in Switzerland) who still use the language in at least some domains (family, school, etc.). Although the number of Romansh speakers has remained surprisingly stable since the first comparable census in 1880, its portion of the population of the canton of Graubünden has decreased from 40% to 14% (main language) and 21% (domain language). Even in some territories previously Romansch dominant, the Romansh speakers have nowadays become a minority. Sutselva and Upper Engadine are in fact falling out of use.

The language territory of Dolomitic Ladin is nowadays divided among three Italian provinces: Bolzano (Gadera Valley, Gardena), Trentino (Fassa), and Belluno (Ampezzo, Livinallongo). The first known text dates from 1631, but until the end of the 19th century very little was written in Ladin and only a few books were published. Towards the end of the 19th century, Ladin was used more frequently in writing, but between 1915 and 1948 the efforts for the written use of Ladin were restrained for political reasons. Fragmentation occurred between 1923 and 1927

under fascist rule and was intended to weaken Ladin, accompanied by other measures to suppress this language. In 1948, Ladin was given special status in Bolzano and Trentino. The Ladins of Ampezzo and Livinallongo (Belluno) did not receive a similar Special Statute.

In connection with the introduction of Ladin in schools and partly also in the administration, different regional written standards were developed, first for the Gardena Valley (Gardenese). This written standard was first also used in the Gader Valley, which later (1970) went its own way (Badiot). Somewhat later, a regional written standard was also developed in the Fassa Valley (Fassan). It is taught in school too, though only the basics. Also for Ampezzo (Ampezzan) and Livinallongo (Fodom), regional written standards were developed. Their use is restricted, as in these regions Ladin is not taught in schools. Since 1988, there have been efforts to develop a uniform written standard for all Ladin speakers, the *Ladin Dolomitan*. The dialectal differences are not as big as in Swiss Romansh; the territory of Dolomitic Ladin is much more compact than the Swiss Romansh territory, although topographic barriers also make contacts difficult.

There's no reliable data on the number of speakers of Ladin. They are included in official statistics only in the province Bolzano; in the other territories they still count as speakers of Italian. In the province Bolzano, 18 736 inhabitants consider themselves as Ladins and in eight communities, the Ladins still have large majorities of between 82% and 97% (2001). With regard to the population of the province of South Tyrol, however, this amount represents only slightly more than 4%. According to estimates, the Ladins are also a majority in the other communities in their territory, with exception of Ampezzo. It is estimated that the total of Ladin speakers amounts to 27 000–30 000.

The Comelico and Cadorine dialects are sometimes associated with Ladin. Recent dialectometrical research, however, does not seem to confirm this affiliation.

The language area of Friulian is situated in the northeast of Italy, in the provinces Pordenone, Udine, and Gorizia, and reaches from Forni Avoltri to just outside Trieste. In the west, the borders of the linguistic frontier to the Venetian dialects are fluid; a transitional zone can be discerned. Already in 1336, the language appears in a coherent text and in the second half of the 14th century, the texts become more frequent. The two oldest poems also date from the 14th century. In Friuli, the art of poetry starts at the beginning of the 16th century and already in one of the first sonnets, *In laude de lenghe furlane*

'In praise of the Friulian language' by Girolamo Sini, writing in Friulian is a topic. This first great age of Friulian literature came to a halt under the influence of Venetian; from 1420 to 1797 Friuli was under Venetian rule. In the 19th century, however, the literary activity was taken up again and has continued until today. Contrary to Swiss Romansh, this re-interest has not led to the development of regional written standards, although Friulian dialectal regions exist: western Friulian, central Friulian, also called east-central Friulian, and Carnic Friulian in the north. Since 1963, Friuli is part of the autonomous region Friuli-Venezia Giulia.

The dialectal differences in Friulian are, however, smaller than in Dolomitic Ladin and in Swiss Romansh, and the mutual understanding assured at all time. There are, however, certain tendencies to use one supra-regional variety in literature, namely the central Friulian of the region around Udine, also called 'common Friulian.' These tendencies, however, are not significant and have been contradicted by opposing tendencies reconsidering the use of the local dialect also as literary language. Since 1985 systematic efforts have been made to create a uniform Friulian written standard. These became particularly urgent as efforts were made to use Friulian in school, on a voluntary basis though, which was rendered possible by the minority law in 1999. Based on this use of the language in a new domain an official orthography was sanctioned the same year and seems to be well on the way.

The number of speakers of Friulian can once again only be based on estimates, as it has not been recorded statistically either. In the actual language territory, it has been estimated at about 600 000, to which 300 000 from outside the territory should be added. The total of speakers of Friulian in the region Friuli-Venezia Giulia would thus amount to about 50%. In any case, Friulian certainly has a sufficiently large number of speakers not to be considered an endangered language. The actual use of the language, however, still seems to be limited to a great extent to the family. It is also to be considered that Friulian has to compete with another Romance language, Italian, contrary to Romansh in Graubünden and Ladin in the Dolomitic Alps, which live in competition with German.

The summing up of these three linguistic areas under the term 'Rhaeto-Romance,' which happened in the second part of the 19th century, was mainly based on some phonetical (e.g., maintenance of *Cl-*) and morphological (e.g., maintenance of *-s* in the plural and in the 2.sg.pl.) particularities that mainly distinguish these linguistic areas from the adjacent northern Italian dialects. This differentiation was

questioned beginning in 1912 (Salvioni) in connection with the changing political situation and the rise of the national states, first and foremost by Italian scholars. The scientific part of the discussion was mainly about the importance of the common features of the three speech communities. These common features were then contrasted with similar features in the neighboring Italian dialects. In the 1980s, thanks to the newly developing dialectometry, this discussion has been resumed on a scientifically more solid basis. Electronic data processing of enormous amounts of material allowed comparison, and it became possible to show that the Rhaeto-Romance linguistic area, based on the maps of the *Dialect Atlas of Italy and Southern Switzerland* (AIS). At the time of the recording of the AIS, northern Italian dialects clearly differed from each other on the one side, but on the other side, they presented significantly more common features (Goebel, 1984). Friulian, however, differed less with its surrounding northern Italian dialects than Swiss Romansh or Dolomitic Ladin. The results of the AIS have recently been confirmed and refined with data, collected since 1985, of the *Atlant linguistisch dl ladin dolomitich y di dialec vejins* (ALD). The detailed interpretation of these results, however, is pending.

In any case, since the 17th century there are no direct frontiers between these three linguistic areas anymore. It is to be noted, however, that all three language areas use their own regional written standards, even though these can by far not cover all written domains, which strongly distinguishes

them from the bordering northern Italian dialect regions.

Bibliography

- Bauer R (2003). *Dialektometrische Analyse des Sprachatlases des Dolomitenladinischen und angrenzender Dialekte* (ALD-I) (2 vols). Salzburg: Institut für Romanistik.
- Billigmeier R H (1979). *A crisis in Swiss pluralism: the Romansh and their relations with the German- and Italian-Swiss in the perspective of a millennium*. The Hague/Paris: Mouton.
- Decurtins A (1965). 'Das Rätoromanische und die Sprachforschung.' *Vox Romanica* 23, 256–304.
- Goebel H (1984). *Dialektometrische Studien: anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF* (3 vols). Tübingen: Niemeyer.
- Gregor D B (1975). *Friulan, language and literature*. New York: Oleander Press.
- Gross M (2004). *Romansh: facts and figures*. Chur: Lia Rumantscha.
- Haiman J & Benincà P (1992). *The Rhaeto-Romance languages*. London/New York: Routledge.
- Holtus G, Metzeltin M & Schmitt C (eds.) (1989). *Lexikon der Romanistischen Linguistik* 3. Tübingen: Niemeyer.
- Heinemann S (2003). *Studien zur Stellung des Friaulischen in der nördlichen Italo-romania*. Bonn: Romanistischer Verlag.
- Kristol A M (1998). 'Die historische Klassifikation der Romania III, Rätoromanisch.' In Holtus G, Metzeltin M & Schmitt C (eds.) *Lexikon der Romanistischen Linguistik* 7. Tübingen: Niemeyer. 937–948.
- Liver R (1999). *Rätoromanisch, Eine Einführung in das Bündnerromanische*. Tübingen: Gunter Narr.

Riau Indonesian

D Gil, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

© 2006 Elsevier Ltd. All rights reserved.

Riau Indonesian is the variety of colloquial Indonesian spoken by the inhabitants of the Indonesian province of Riau, which encompasses parts of east-central Sumatra plus a large number of adjacent smaller islands. Riau Indonesian is one of many distinct local varieties of colloquial Indonesian spoken throughout the archipelago, such as, for example, Jakarta Indonesian. The population of Riau province, numbering close to 5 million people, is linguistically and ethnically heterogeneous. Although the indigenous population is mostly Malay, a majority of the

present-day inhabitants are migrants from other provinces, speaking a variety of other languages. Riau Indonesian is acquired as a native language by most or all children growing up in Riau province, whatever their ethnicity. It is the language most commonly used as a lingua franca for interethnic communication, and, in addition, it is gradually replacing other languages and dialects as a vehicle for intraethnic communication.

Riau Indonesian is quite different from Standard Indonesian, a language familiar to many general linguists from a substantial descriptive and theoretical literature. Riau Indonesian is also distinct from a set of dialects generally referred to as Riau Malay, used in Riau province by ethnic Malays, primarily for intraethnic communication. In addition, Riau

Indonesian is distinguished from another set of Malayic dialects spoken by various indigenous peoples in Riau province (Orang Asli, Orang Sakai, Orang Akit, Orang Hutan, and Orang Laut). Finally, Riau Indonesian is also different from the variety of Malay/Indonesian sometimes referred to as 'Bazaar Malay,' which is used by the ethnic Chinese residents of Riau province when speaking to non-Chinese, and by the non-Chinese when speaking to the ethnic Chinese. Thus, the sociolinguistic situation in Riau province is one of great complexity: speakers of Riau Indonesian are often fluent in several other varieties of Malay/Indonesian, as well as in other languages, such as Minangkabau and Javanese.

From a general typological perspective, Riau Indonesian is a strongly isolating language, with no inflectional morphology and relatively little derivational morphology or compounding. It is also a language with very flexible word order. Perhaps the most striking feature of Riau Indonesian is the pervasiveness of underspecification, i.e., the absence of obligatory overt grammatical expression for a wide variety of semantic categories, including number, definiteness,

tense, aspect, thematic role, and ontological type. On the basis of these characteristics, Riau Indonesian has been argued to have a simple grammar, lacking much of the machinery central to most grammatical theories. Syntactically, it is said to have a single open syntactic category, that is to say, no distinction between nouns, verbs, adjectives, and prepositions, or between lexical categories and phrasal ones. Semantically, it is claimed that when two or more expressions are combined, the meaning of the combination is usually associated with the meanings of the constituents in a vague and underspecified fashion.

Bibliography

- Gil D (1994). 'The structure of Riau Indonesian.' *Nordic Journal of Linguistics* 17, 179–200.
 Gil D (2001). 'Creoles, complexity and Riau Indonesian.' *Linguistic Typology* 5, 325–371.
 Gil D (2005). 'Isolating-monocategorial-associational language.' In Cohen H & Lefebvre C (eds.) *Categorization in cognitive science*. Oxford: Elsevier.

Romance Languages

R Wright, University of Liverpool, Liverpool, UK

© 2006 Elsevier Ltd. All rights reserved.

The Romance languages are those that have developed from the spoken Latin of the early Middle Ages. In this sense one can claim that Latin is not dead; about a quarter of the world's population still speak it; but it has acquired several new geographically based names (Spanish, Portuguese, French, Italian, Romanian, Catalan, Occitan, Sardinian, Galician, Rhaeto-romanian, here listed roughly in descending order of number of speakers). These are for political reasons considered to be separate 'Romance' languages, but there is still essentially one dialect continuum overlaid by the several artificially extended standards. Apart perhaps from Romanian, the location and history of whose earliest speakers is still controversial, the definitive divergence into separately identifiable languages should be dated to no earlier than the 9th century, and in several cases later.

Reconstruction

There are two main kinds of evidence for the Romance (spoken Latin) that existed before the

separate languages diverged: surviving written texts and the results of 'reconstruction.' Hall and others have reconstructed a hypothetical 'Proto-Romance' on the basis of the later Romance languages; features they have in common are taken to have existed in their ancestors. As compared with 'classical' Latin, this Proto-Romance contains, for example, no neuter nouns, no ablative cases, no datives and genitives outside pronouns, no synthetic passives or futures, no adverbs in *-iter*, no phonemic length distinctions in vowels, no originally final consonants other than alveolars, and no velar consonants before front vowels other than those that were originally labio-velar. On the other hand, the evidence of modern Romance languages suggests that their base included extended uses of prepositions (particularly *ad* and *de* to replace inflectional nominal suffixes); analytic passives with auxiliary *esse* + tense-indeterminate participles; extended use of grammatically reflexive *se* with passive meaning; analytic futures (and 'conditionals') formed with the infinitive + *habeo*; new analytic perfects (including future perfects and pluperfects) formed with activated participles + *habeo*; extensive use of *ille* and *ipse* with the functions of the definite article; many diminutives in *-iculum* and other affixed forms (such as *adiutare*,

rather than *iuuare*, as the base of Port *ajudar*, Sp *ayudar*, Cat *ajudar*, Fre *aider*, Italian *aiutare*, Romanian *ajuta*, etc.); the use of preposed *magis* or *plus* instead of comparative *-ior*; palatal affricates and semivowels; and much new vocabulary from, in particular, Germanic sources.

Texts

This reconstructed language is not very much like that of the surviving written texts of the time. Janson described reconstruction and textual analysis as two different key-holes through which one can look into the same large room, for the rules of correct writing did not change, and ‘mistakes’ are rarely attested. Most texts were written on perishable wax tablets or papyrus; the extant versions are usually later manuscript copies prepared by scribes who had specific instructions to ‘correct’ their originals according to the arcane and eventually archaic rigidities of the Imperial grammarians. Texts without such distortions are few; Adams has published some letters and drafts, and Väänänen’s study of the Pompeii *Graffiti* revolutionized the discipline by showing how ‘incorrectly’ nonscholars wrote in 79 A.D. Even these texts, however, are obviously not phonetic transcriptions of actual speech. From painstaking statistical analyses of surviving inscriptions (mostly on tombstones), whose textual details cannot be ‘corrected,’ Herman has concluded that Imperial spoken Latin was evolving but also converging, with new features starting in one place becoming eventually attested anywhere. Some further progress is made by studying borrowings from spoken Latin into, for example, Irish, Welsh, Berber, Albanian, and Greek.

Divergence

Wide variation arose, but this need not imply mutual unintelligibility. Many historians, textual critics, philologists, sociolinguists, and historical linguists currently view early Medieval Romance Europe as a single lively speech community, where almost everyone could understand old-fashioned written texts when read aloud (McKitterick, 1990, Wright, 1991). These were not ‘Dark’ Ages. Early Medieval speakers rarely made metalinguistic distinctions that we take for granted now, neither diatopic (between French, Spanish, etc.) nor diastratic (between Romance and Medieval Latin). The latter distinction was probably imported from Germanic-speaking areas, where

vernacular Germanic and official Latin *grammatica* were unrelated and self-evidently different languages; conscious distinctions between separate Romance languages became widespread only after the fashion for inventing distinctive writing systems in different areas, which began experimentally in 9th-century eastern France but generalized only in the 12th and 13th centuries. Indeed, to some extent the speech of the central Romance area is still mutually intelligible, given goodwill and clarity from those in the conversation; peripheral languages, such as Romanian, Portuguese, and French, are rarely intelligible elsewhere.

Bibliography

- Adams J N (1977). *The vulgar Latin of the letters of Claudius Terentianus*. Manchester: Manchester University Press.
- Adams J N (2003). *Bilingualism and the Latin language*. Oxford: Oxford University Press.
- Banniard M (1992). *Viva Voce: communication écrite et communication orale du IVe au IXe siècle en Occident latin*. Paris: Institut des Études Augustiniennes.
- Banniard M (ed.) (2002). *Langages et peuples d'Europe*. Toulouse-le-Mirail: CNRS.
- Hall R A (1976). *Proto-Romance phonology*. New York: Elsevier.
- Harris M & Vincent N (eds.) (1987). *The Romance languages*. London: Croom Helm.
- Herman J (1990). *Du latin aux langues romanes. Études de linguistique historique*. Tübingen: Niemeyer.
- Herman J (ed.) (1998). *La transizione dal latino alle lingue romanze*. Tübingen: Niemeyer.
- Herman J (2000). *Vulgar Latin*. Pennsylvania: Penn State Press.
- Janson T (1979). *Mechanisms of language change in Latin*. Stockholm: Almqvist and Wiksell.
- McKitterick R (ed.) (1990). *The uses of literacy in early Medieval Europe*. Cambridge: Cambridge University Press.
- Selig M (ed.) (1993). *Le passage à l'écrit des langues romanes*. Tübingen: Narr.
- Trotter D A & Gregory S (eds.) (1997). *De mot en mot: aspects of medieval linguistics*. Cardiff: University of Wales Press.
- Väänänen V (1937). *Le latin vulgaire des inscriptions pompeïennes*. Helsinki: Academy.
- Wright R (ed.) (1991). *Latin and the Romance languages in the early Middle Ages*. London: Routledge.
- Wright R (1995). *Early Ibero-Romance*. Newark, DE: Juan de la Cuesta.
- Wright R (2003). *A sociophilological study of late Latin*. Turnhout: Brepols.

Romani

Y Matras, University of Manchester, Manchester, UK

© 2006 Elsevier Ltd. All rights reserved.

Definitions

Romani (referred to by its speakers as *řomani čhib* ‘the Romani language’ or *řomanes* ‘in a Romani way’) is the only Indo-Aryan language spoken exclusively in Europe, as well as by emigrant populations in the Americas and Australia. The language is often referred to as ‘Gypsy’; it is important, however, to distinguish between Romani, which is the fully fledged, everyday family and community language spoken by the people who call themselves *Řom*, and secret or in-group vocabularies employed in various parts of the world, including in Europe, by other populations of peripatetics or so-called service-nomads. There is nevertheless some interface between the two phenomena: in some regions of Europe, especially the western margins (Britain, the Iberian peninsula, Scandinavia), Romani-speaking communities have given up their language in favor of the majority language but have retained Romani-derived vocabulary as an in-group code. Such codes, for instance *Angloromani* (Britain), *Caló* (Spain), or *Rommani* (Scandinavia) are usually referred to as *Para-Romani* varieties.

In the absence of reliable census figures, the total population of Romani speakers can only be estimated, at anywhere upwards of 3.5 million. The largest concentrations of Romani speakers are in southeastern and central Europe, especially Macedonia, Bulgaria, Romania, and Slovakia. Romani has traditionally been an oral language, and in more traditional communities there is even opposition to codification attempts or other public use of the language, which is viewed as having protective functions. The overwhelming trend, however, since the early 1990s has been toward codification of the various dialects at local or regional levels. The language is now used in local media, on numerous Internet sites, as a medium of correspondence (especially electronic), and in some countries even as a medium of school instruction.

History

The earliest attestation of Romani is from 1542, in Western Europe. Our understanding of the language’s historical development is therefore dependent on reconstruction and comparison with other Indo-Aryan idioms as well as with the contact languages. In phonology, Romani shares a number of ancient isoglosses with the Central branch of Indo-Aryan, most notably

the realization of Old Indo-Aryan *r* as *u* or *i* (Sanskrit *śṛṇ-* Romani *šun-* ‘to hear’) and of *kṣ-* as *kh* (Sanskrit *akṣi* Romani *j-akh* ‘eye’). In contrast, however, to the other Central languages, Romani preserves a number of dental clusters (Romani *trin* ‘three’, *phral* ‘brother’; cf. Hindi *tīn*, *bhāī*). This led Turner (1926) to assume a Central origin of Romani, with subsequent migration to the Northwest before the reduction of the relevant clusters took place. A north-western migration is of course well in line with an ultimate migration out of India and on towards Europe. Further support for Turner’s theory comes from the domain of verb morphology, where Romani follows the exact same pattern as Northwestern languages such as Kashmiri or Shina in its renewal of the past-tense conjugation through the adoption of oblique enclitic pronouns as person markers (*kerdo* ‘done’ + *me* ‘me’ > *kerdjom* ‘I did’). Proto- or pre-European Romani was thus a kind of Indian hybrid: a central Indic dialect that had undergone partial convergence with northern Indic languages. Although the retention of dental clusters would suggest a break with the Central languages during the transition period from Old to Middle Indo-Aryan, the overall morphology of Romani indicates that the language participated in some of the significant developments leading toward the emergence of New Indo-Aryan (such as the reduction of the nominal case system to a two-way opposition, nominative vs. oblique, and grammaticalisation of new, postposed case markers). It would appear therefore that Proto-Romani did not leave the Indian subcontinent until late in the second half of the first millennium CE. Romani is among the most conservative New Indo-Aryan languages in retaining a full consonantal present conjugation, as well as consonantal oblique nominal case endings. Typical phonological developments that characterize Romani among the Indo-Aryan languages are the devoicing of aspirates *bh*, *dh*, *gh* to *ph*, *th*, *kh*, the shift of medial *d*, *t* to *l*, of short *a* to *e*, of inflectional *-a* to *-o*, of initial *kh* to *x*, and of the retroflexes *ḍ*, *ṭ*, *ḍḍ*, *ṭṭ*, *ḍh* etc. to *r* and *ř*.

The subsequent development of the language was strongly influenced by its contact languages. Romani borrowed lexicon and some grammatical vocabulary from Iranian languages and Armenian. The heaviest impact on Early Romani (European Romani, between the 10th and 13th centuries C.E.) was of Byzantine Greek. Apart from numerous lexical loans, phonemes, and grammatical vocabulary, Romani adopted Greek inflectional morphology in nouns and verbs, which remain productive with loan vocabulary from subsequent European contact languages (see below). Greek

also had a strong impact on the syntax of Romani, triggering among other things a shift to VO word order and the emergence of a preposed definite article.

The sound system

Romani dialects generally preserve an aspirated set of voiceless stops *ph*, *th*, *kh* as well as *čh*, alongside *p*, *t*, *k*, *č* and *b*, *d*, *g*, *dž*. Nasals are *m* and *n*, fricatives are *f*, *v*, *x*, *h*, *s*, *z*, *š*, and in some dialects also *ž*, and there is an affricate *c* [ts]. All dialects have *l* and *r*, and some also retain *ř*, which is realized as either a uvular [ʀ], a long trill [rr], or in some dialects a retroflex [ɽ, ɽ̥]. Palatalization of consonants, either distinctive or nondistinctive, is common in the Romani dialects of eastern and southeastern Europe. The vowel system consists of *a*, *e*, *i*, *o*, *u*, with addition in some dialects of a central vowel ə or ɨ. Western European dialects of Romani tend to show vowel length distinctions. The phoneme inventory of individual dialects usually accommodates additional phonemes from the respective contact languages in lexical loans. Conservative stress in Romani is on the final inflectional segment of the word, though a number of affixes remain unstressed, among them the vocative ending, agglutinative (Layer II) case endings (see below), and the remoteness tense marker. Dialects in Western and Central Europe often show a shift of stress to earlier positions in the word.

Morphology

Nominal forms

Romani nominal morphology is inflectional, with some agglutination. There are two genders, masculine and feminine, and two numbers, singular and plural. Mass nouns often allow omission of overt plural marking. The principal inflectional alternation in the noun is between two ‘basic’ or Layer I cases, nominative and oblique, in the singular and plural. The different patterns of alternation constitute declension classes. Romani declension classes are sensitive to gender, to the phonological shape of the stem, and to etymology, with European loans taking Greek-derived case endings. Basic inherited declension classes are presented in Table 1.

Individual dialects show various patterns of analogies among the different classes. Loan declension classes typically have Greek-derived inflection endings *-os*, *-o*, *-is*, or *-us* (masculine) and *-a* (feminine), with a variety of plural endings such as *-i*, *-e*, *-ides*, *-uri* and more. The oblique stem serves as the base for further (Layer II) agglutinative case formation, with the endings *-tel/-de* (locative and prepositional),

Table 1 Basic ikeoclitic declension classes

| | Sg. nominative | Sg. oblique | Pl. nominative | Pl. oblique |
|---------------------|---------------------------|------------------|-------------------|------------------|
| Masculines in –o | <i>čhav-o</i> ‘boy’ | <i>čhav-es-</i> | <i>čhav-e</i> | <i>čhav-en-</i> |
| Masculines in –φ | <i>phral</i> ‘brother’ | <i>phral-es-</i> | <i>phral-a</i> | <i>phral-en</i> |
| Feminines in –i | <i>řomn-i</i> ‘woman’ | <i>řomn-ja-</i> | <i>řomn-ja</i> | <i>řomn-jen-</i> |
| Feminines in –φ | <i>phen</i> ‘sister’ | <i>phen-a-</i> | <i>phen-a</i> | <i>phen-en-</i> |

-kel/-ge (dative), *-tar/-dar* (ablative), *-sa(r)* (instrumental and comitative), and *-ker/-l-ger-* (genitive). As in other Indo-Aryan languages, the genitive agrees with the head noun (*čhav-es-ker-o phral* ‘the boy’s brother,’ *čhav-es-ker-i phen* ‘the boy’s sister’). The oblique without a Layer II extension serves as the case of the direct object (‘accusative’) with animate nouns.

Adjectives usually take vowel endings that agree with the vocalic case-endings of the noun (*mir-o dad* ‘my father,’ *mir-i daj* ‘my mother’). Demonstratives usually show a four-term system, encoding both proximity/remoteness (or, rather, presence in the situation vs. the discourse context), and general/specific (disambiguation), e.g., *adava*, *akava*, *odova*, *okova*. Interrogatives are cognate with other Indo-Aryan languages (*kon* ‘who,’ *kaj* ‘where’), with *so* ‘what’ serving as the base for several derived forms (*savo* ‘which,’ *soske* ‘why,’ *sode* ‘how many,’ etc.). Indefinite markers are often borrowed from the respective contact languages.

Verbs

Valency is a central feature of Romani verb morphology. It is expressed through direct affixation to the verb root. The productivity, however, of individual valency markers varies among the dialects. Typical valency-increasing markers are *-av-*, *-ar-*, *-ker-*, and valency-decreasing markers are *-jov-* and *-áv-*. They derive verbs from other verb roots, as well as from nouns and adjectives. Borrowed verbs carry loan verb extension or adaptation markers, based on Greek-derived tense/aspect affixes such as *-iz-*, *-in-*, *-is-*, sometimes in combination with valency affixes (e.g., *-is-ar-*, *-is-ker-*).

The default stem (root with derivation marker) serves as a non-perfective aspect. The plain form of the nonperfective serves as a present/subjunctive. A tense/modality extension *-a* marks the present/indicative, the future, or conditional, depending on the dialect. A perfective aspect (also ‘aorist’ or ‘simple past’) is formed by attaching a perfective extension (derived from the Middle Indo-Aryan participle extension *-t-*) to the root of the verb

(e.g. *ker-d* ‘did’). The choice of perfective extension depends on the numerous perfective classes, which are sensitive to the root phonology as well as to valency and semantics.

There are two person conjugations: The present conjugation (1sg *-av*, 2sg *-es*, 3sg *-el*, 1pl *-as*, 2/3pl *-en*) continues the Middle Indo-Aryan set of present concord markers. There are two inflection classes in the present (nonperfective), distinguishing vocalic and consonantal roots (*xa-s* ‘you eat,’ *kam-es* ‘you-want’). The perfective conjugation, which follows the perfective extension, derives from late Middle Indo-Aryan enclitic pronouns (1sg *-om*, 2sg *-al/an*, 3sg *-as*, 1pl *-am*, 2pl *-an/-en*, 3pl *-e*).

Both the present and the perfective may be extended by a remoteness marker *-as/-ahi/-ys/-s* that is external to the subject concord marker, indicating the imperfect/habitual/conditional (with the present) or the pluperfect/counterfactual (with the perfective).

Syntax

Romani stands out among the Indic languages through its Europeanized, specifically Balkanized syntax. Word order is VO, with variation between thetic (continuative) VS and categorical (contrastive) SV. Local relations are indicated by prepositions. Adjectives and determiners generally precede the noun, as does the definite article (which agrees with the noun in gender, number, and case). Relative clauses are postposed, and often introduced by a universal relativizer *kaj* < ‘where.’ Clauses are generally finite. Adverbial clauses are introduced by conjunctions, usually derived from interrogatives. Romani distinguishes between factual and nonfactual complex clauses. Modal, manipulation, and purpose clauses are introduced by a nonfactual conjunction *te*, as are conditional clauses. Epistemic complements

are introduced by *kaj*, which is often replaced by a borrowing from the respective contact language.

Dialect diversity

Dialect differentiation in Romani appears to have emerged largely in situ, following the dispersal of groups from the Balkans into western and northern Europe, from around the 14th century onward, and their settlement in their present locations, during the 16th–17th centuries. There are two major diffusion centers of innovations: in the southeast, especially the northern Balkans, and in western-central Europe, especially Germany. Typical of the western-northern dialects are prothesis of *j*-, simplification of *ndř* to *r*, loss of adjectival past-tense in intransitives (*gelo*, *geli* > *geljas* ‘he/she went’), and retention of *-n* in the abstract nominalizer *-ipen/-iben*. In the central regions, *s* in grammatical paradigms is often replaced by *h*. Individual regions show distinct developments in morphological paradigms, especially demonstratives, 2/3pl perfective concord markers, and loan verb markers. Especially these latter isoglosses justify the current classification into the following dialect groups: Balkan (with a subgroup ‘Black Sea Coast’), Vlax (Transylvannia and adjoining regions), Central, Northeast (Baltic-Norussian), and Northwest (German-Scandinavian).

Bibliography

- Bakker P & Matras Y (2003). *Bibliography of modern Romani linguistics*. Amsterdam: John Benjamins.
 Elšik V & Matras Y (eds.) (2000). *Grammatical relations in Romani: the noun phrase*. Amsterdam: John Benjamins.
 Matras Y (2002). *Romani: a linguistic introduction*. Cambridge: Cambridge University Press.
 Turner R (1926). ‘The position of Romani in Indo-Aryan.’ *Journal of the Gypsy Lore Society* 5(3), 145–189.

Romanian

J Augerot, University of Washington, Seattle, WA, USA

© 2006 Elsevier Ltd. All rights reserved.

The Romanian (alternatively spelled Rumanian, Roumanian) language is a member of the Romance languages and is a continuation of Eastern Latin, spoken in the Roman province of Dacia and surrounding areas which were colonized by the Romans led by the emperor Trajan from 106 A.D.

The Eastern Latin spoken in this area by Roman soldiers and colonists along with presumably assimilated local tribes underwent numerous influences after the Roman legions abandoned the northernmost areas in 271 A.D. The first major influence was undoubtedly provided by the indigenous Dacian–Thracian–Illyrian inhabitants of the area, although only a few undisputed words from that source remain. A. Rosetti mentions some 88 Indo-European but pre-Latin terms whose existence in both Romanian and Albanian suggests a

substratum as the source for these words in both languages, e.g., *barză* 'stork,' *bucurie* 'happiness,' *mănz* 'colt,' *mos* 'old man'; cf. Albanian *bardhë* 'white (FEM),' *bukuri* 'beauty,' *mëz* 'colt,' *moshë* '(old) age.' It is likely that some of Romanian's morphosyntactic structure, especially that shared with the non-Romance Balkan languages, also comes from this pre-Latin substratum, but the complete lack of textual evidence leaves these possibilities in the realm of speculation.

The longest and most consistent influence was that of the Slavs, who began to migrate across this territory to the south as early as 400 A.D. and who remain neighbors on three sides of present-day Romania. Several other migrations of peoples crossed the territory of today's Romania between the departure of the Roman legions and the appearance of the modern Romanian people, but little or no linguistic traces of the Goths, Huns, Gepids, Avars, *et al.* remain.

This early Romanian soon (perhaps as early as the 10th century) began to split, first into four dialects which later tended to become languages in their own right: the principal one in terms of numbers is termed Daco-Romanian, spoken primarily north of the Danube. The second largest division is called Aromanian (Macedo Romanian) and is currently spoken in pockets of southwestern Bulgaria, Macedonia, Albania, and northern Greece. The other two are quite limited in extent: Megleno Romanian is confined to an area in southeastern Macedonia and adjacent northern Greece, and Istro Romanian is limited to eight localities in Istria in modern Croatia.

Daco-Romanian is now called Romanian and is the single national language among the four with approximately 25 million speakers worldwide. It shows, in addition to the Slavic mentioned above, the influences of prolonged contact with Hungarian, Turkish, and early modern Greek. More recently, from the 19th century onwards, Romanian has been subjected to a kind of re-Latinization, both from the strong influence of French and from the international European vocabulary largely based on Latin and Italian.

The first preserved texts in Romanian date from the end of the 15th and beginning of the 16th centuries. The oldest dated text (1521) is a letter to Neacșu of Câmpulung.

Although some texts from as early as the 16th century are written in the Latin alphabet, which is the norm today, until 1860 the official alphabet was Cyrillic. Most of the letters have approximately their general European values, but there is some use of diacritics that should be noted, namely *î*, *â*, *ă*, *ț*, *ș*. Phonetic equivalents are given where necessary in brackets in Tables 1 and 2.

Table 1 Vowels

| | Front | Central | Back rounded |
|------|-------|-----------|--------------|
| High | i | î (â) [i] | u |
| Mid | e | ă [ə] | o |
| Low | | a | |

Table 2 Consonants

| | Bilabial | Labio-dental | Dental | Palato-alveolar | Dorso-velar |
|------------|----------|--------------|--------|-----------------|-------------|
| Stops | p b | | t d | ch [k'] gh [g'] | c [k] g |
| Affricates | | | ț [c] | c [č] g [č] | |
| Fricatives | | f v | s z | ș [š] j [ž] | h |
| Nasals | m | | n | | |
| Lateral | | | l | | |
| Trill | | | r | | |

Romanian symbolizes diphthongs with pairs of vowels, e.g., *iu*, *ia*, *ea*, *ua*, *oa* and *ai*, *au*, *ou*, *ui*, but one can seldom be certain from the spelling alone. See the website of the Romanian Academy Center for Artificial Intelligence for a complete inventory of multiple vowel combinations.

In contrast to the practice of some of its immediate neighbors, there is no final devoicing of consonants as in Bulgarian and Russian and no contrastive vowel length as in Slovak, Hungarian, and Serbian (Serbo-Croatian). Rather, Romanian forms a bridge between Serbian and Ukrainian in terms of final voiced obstruents.

The two palatal velars are disputed as separate from the dorsovelars in phonology but are clearly different on the morphophonemic plane. They represent an interesting progression from Latin 'cl' and 'gl' in words such as *cheie* 'key' and *gheață* 'ice.' Intentionally omitted from the chart are the palatalized variants of most consonants that, according to some analyses, occur word finally before the final (voiceless?) [i], for example in the second person singular present forms of the verb, e.g., *tu plimbi* 'you stroll,' *vezi* 'you see,' *scoli* 'you arise.'

Word stress in Romanian is basically penultimate, but only if one counts final (mostly silent and, in the case of /u/, mostly unwritten) /i/, /u/ as a syllable (the stressed vowel is given in bold type): *plecare-pleacă-plecând*, *venire-vine-venind*, *miere*. There are several exceptions to this generalization including many borrowings, e.g., through Turkish: *baklava*, *sarma*, *cafea*, and the majority of the infinitive forms, which are also stressed on the final syllable: *veni*, *citi*, *pleca*, *turna*, *putea* (historically, of course, the infinitive had

Table 3 Morphophonemics

| underlying forms | mes-ə mes-e | fet-ə fet-e | mer-u mer-e | per-u per-i |
|----------------------|-------------|-------------|-------------|-------------|
| 1. breaking | m̥easə -- | f̥eata -- | -- -- | -- -- |
| 2. backing | m̥easə -- | f̥ata | m̥aru -- | p̥aru -- |
| 3. coalescence | masə mese | fata fete | -- -- | -- -- |
| 4. drop final /u, i/ | -- -- | -- -- | m̥ar mere | p̥ar per' |
| normal spelling | masă mese | fată fete | măr mere | păr peri |

a final syllable ‘-re’). Other exceptions include the classic Latin deviations based on heavy syllables as in *industrie* and words of specific morphological categories with stress fixed on a certain ending, such as the imperfect tense: *plecam*, *plecai*, *pleca*, etc. Note that the addition of the masculine genitive definite article /-u-lui/, the feminine definite plural article /-i-le/ or the ‘multisyllabic’ plural desinence /-uri/ does not cause movement of the stress: *marfă*, *mărfuri*. Thus it is not truly penultimate, only dominantly so, since one must proceed from a hypothetical base form of the word and include several classes of exceptions.

From a morphophonemic point of view, Romanian is quite complex. Space limitations will not allow a complete treatment here, but an illustration of this complexity may be seen in the singular and plural forms of words such as *masă*, *mese* ‘table(s)’, *fată*, *fete* ‘girl(s)’, *măr*, *mere* ‘apple(s)’, *păr*, *peri* ‘hair(s)’, where only by beginning with a root with the front vowel /e/ may one predict all the forms of any one word. The three rules that are involved are illustrated in Table 3.

Rule 4 brings up some other complications such as the phonetic interpretation of consonants preceding the final ‘dropped’ /-i/, but it can be seen that, at least for inherited Latin vocabulary, these rules will predict the surface phonetics of many forms. Compare (given here in the normal orthography) the various forms of the verbs *a vedea* ‘to see’ and *a învăța* ‘to learn’ with the same stressed root vowel /e/, i.e. /véd-/ , /învéc-/ (see Table 4).

It can be seen that the same rules allow the derivation of all the verbal forms if we add the ‘flip-flop’ rule, namely, in the conjunctive-subjunctive the present third singular ending /-e/ goes to /-ə/ and vice versa. Thus *el pleacă* ‘he leaves’ becomes *el o să plece* while *el crede* ‘he believes’ becomes *el o să creadă*.

It should be noted that the stressed schwa is a feature that is often included among the features of the so-called Balkan Sprachbund. There are characteristics (merging of the genitive and dative, tendency to lose the infinitive, development of a postpositive definite article, repetition pronominally of the direct

Table 4 Verbal forms with root in /e/

| | | |
|-----------------------|---------|-----------|
| present 1st sing. | văd | învăț |
| present 2nd sing. | vezi | înveți |
| present 3rd sing. | vede | învăță |
| conjunctive 3rd sing. | să vadă | să învețe |

Table 5 The definite article

| Feminine | Masculine | Neuter |
|------------------------|----------------------------|-------------------------|
| o doamnă <i>a lady</i> | un domn <i>a gentleman</i> | un os <i>a bone</i> |
| doamna doamnele | domnul domnii | osul oasele |
| o masă <i>a table</i> | un pom <i>a tree</i> | un lucru <i>a thing</i> |
| masa mesele | pomul pomii | lucruri lucrurile |
| o carte <i>a book</i> | un perete <i>a wall</i> | un scaun <i>a chair</i> |
| cartea cărțile | peretele pereții | scaunul scaunele |

object) in several areas of the grammar that Romanian shares with languages clearly not of Romance origin such as Bulgarian, Macedonian, Albanian, and dialectal areas of Serbian and modern Greek.

Most noticeable in differentiating Romanian from the rest of Romance is the postposed definite article, which takes several forms but is basically singular *-a*, plural *-le* for feminines and singular *-l*, plural *-i* for masculines (with a special category of neuter that generally takes the masculine article in the singular and the feminine in the plural; examples in Table 5). In the noun phrase the same definite article is attached to the first element in the phrase: *un prim pas* ‘a first step,’ *primul pas* ‘the first step.’

Unlike the rest of Romance, although they have been whittled down to just two (plus a vocative that has almost disappeared), Romanian still maintains nominal case distinctions. Among Romanian scholars there is a tendency to name them nominative-accusative and genitive-dative after the functions of each case; see Table 6 for some examples.

Table 7 gives two verbs illustrating the basic verbal categories: *a intra* ‘to enter’ and *a face* ‘to do.’ There are, in addition to the present tense, two ways to form the future, and a variety of past tenses, including

Table 6 Cases

| Gender | Article | Nom-Acc Singular | Gen-Dat Singular | Nom-Acc Plural | Gen-Dat Plural |
|-----------|------------|-------------------|------------------|----------------|----------------|
| Feminine | Indefinite | casă <i>house</i> | case | case | case |
| | Definite | casa | casei | casele | caselor |
| Masculine | Indefinite | bărbat <i>man</i> | bărbat | bărbați | bărbați |
| | Definite | bărbatul | bărbatului | bărbații | bărbaților |
| Neuter | Indefinite | lemn <i>wood</i> | lemn | lemne | lemne |
| | Definite | lemnul | lemnului | lemnule | lemnelor |

Table 7 Verb tenses

| | 1st sing | 2nd | 3rd | 1st plur | 2nd | 3rd |
|----------------|----------|-----------|---------|------------|-------------|-----------|
| Present | intru | intri | intră | intrăm | intrați | intră |
| | fac | faci | face | facem | faceți | fac |
| Imperfect | intram | intrai | intra | intrăm | intrați | intrau |
| | faceam | faceai | facea | faceam | faceați | faceau |
| Simple perfect | intrai | intrași | intră | intrarăm | intrarăți | intrară |
| | făcui | făcuși | făcu | făcurăm | făcurăți | făcură |
| Pluperfect | intrasem | intraseși | intrase | intraserăm | intraserăți | intraseră |
| | făcusem | făcuseși | făcuse | făcuserăm | făcuserăți | făcuseră |

the compound past, the imperfect, the perfect and the pluperfect. Again it is instructive to consider the dropped final high vowels /u/, /i/ as morphological formants of the first and second persons singular in the present tense, both for purposes of word accent and for various other phonological alternations. The forms are given in standard orthography.

There are two general types of regular verb, one with base + back vowel and the other with base + plus front vowel, and with this information plus sets of personal endings all of the conjugations of a given verb may be produced. I chose here the verb *a intra* since it gives justification for the otherwise phantom ending /-u/ of the first person singular for a-verbs and both first singular and third plural for e-verbs (also true for i-verbs, which are not illustrated here).

All other verbal formations (except for the conjunctive-subjunctive future with *o să* plus the present, discussed above with the flip-flop rule) are synthetic, relying on either the short infinitive or the past passive participle: thus the future is *voi, vei, va, vom, veți, vor* plus the infinitive (*voi intra, vei intra*, etc.), and the conditional is *aș, ai, ar, am, ați, ar* also with the infinitive (*aș intra, ai intra*, etc.); while the compound past uses *am, ai, a, am, ați, au* plus the past passive participial form (*am intrat, ai intrat*, etc.), and the past conditional also uses it after the same auxiliary as the conditional, adding the verb *fi* ‘to be’ (*aș fi intrat, ai fi intrat*, etc.). One further verb form is the past subjunctive, which does not

conjugate and has a single fixed auxiliary *să fi* with the past passive participle (*să fi intrat* for all persons). There is also a ‘future in the past’ with *fi*.

Bibliography

- Agard F (1958). *Structural sketch of Rumanian*. Baltimore: Waverly Press.
- Augerot J E (1974). *Romanian phonology*. Moscow, ID: Idaho Research Foundation.
- Alexandrescu S (ed.) (1977). *Transformational grammar and the Rumanian language*. Lisse, Netherlands: The Peter de Ritter Press.
- Chitoran I (2002). *The phonology of Romanian: a constraint-based approach*. New York: Mouton de Gruyter.
- Coteanu I & Dănăilă I (1970). *Introducere în lingvistica și filologia românească*. Bucharest: Editura Academiei Republicii Socialiste România.
- Petrovici E (1934). ‘Le pseudo *i* final du roumain.’ *Bulletin linguistique* 2, 86.
- Pop L & Moldovan V. *Grammaire du roumain/Romanian grammar*. Cluj: Echinor.
- Pușcariu S (1940). *Limba română*. Bucharest: Fundația Pentru Literatură și Artă.
- Rosetti A (1964–1966). *Istoria limbii române* (6 vols). Bucharest: Editura Științifică.
- Rosetti A (1973). *Brève histoire de la langue roumaine des origines à nos jours*. The Hague: Mouton.
- Sala M (1976). *Contributions à la phonétique historique du roumain*. Paris: Klincksiek.
- Stan I (1996). *Studii de fonetică și fonologie*. Cluj-Napoca: Presa Universitară Clujeana.

Vasiliu E (1965). *Fonologia limbii române*. Bucharest: Editura Științifică.

Vasiliu E & Golopenția Eretescu S (1972). *The transformational syntax of Romanian*. The Hague: Mouton.

Relevant Website

<http://www.racai.ro> – Romanian Academy Center for Artificial Intelligence.

Russenorsk

E H Jahr, Agder University, Kristiansand, Norway

© 2006 Elsevier Ltd. All rights reserved.

The Arctic pidgin Russenorsk (RN), which developed from Norwegian and Russian during the second half of the 18th century, was used in barter trade in northern Norway for around 150 years. RN's main period of use was the 19th century when trading reached large proportions. The sociolinguistic situation in northern Norway during the 19th century was multifaceted and complex, involving many different languages.

RN, now extinct, exhibits several features that make it theoretically interesting. In spite of being formed as a dual-source pidgin – from two Indo-European languages – it shows most of the features common to all pidgin languages. To a stabilized grammatical and lexical core, a variety of lexical items could be added when the situation called for it. It also is noteworthy that RN, unlike most pidgins, was created by socially equal groups of speakers.

Until around 1850, RN enjoyed a high social status, as both fishermen and merchants had to use the pidgin when dealing with the Russians. After 1850, the use of RN was restricted mostly to common fishermen, because the merchants – who constituted the local upper classes – began to spend longer periods of time with colleagues in northern Russia and subsequently developed their own grammatically simplified variety of Russian. As a consequence, RN's social status was devalued.

Today, we have to rely on written material in our study of RN. However, the available texts allow for studies of both lexicon and grammar. They consist of isolated sentences, word lists, and conversations in dialogue form. Altogether, they include some 400 different words, with a core of 150–200 lexemes. Most of them are related to the barter trade. This trade constituted the socioeconomic basis for RN, and when it gradually gave way to cash trade early in the 20th century, the language lost ground and disappeared.

The characteristic features of RN can be summarized as follows:

- a. The phonology reflects Norwegian and Russian – however, sounds and consonant clusters not found in both languages are avoided or simplified.
- b. 1st and 2nd personal/possessive pronouns are *moja* and *tvoja*.
- c. *po* 'on' is the only preposition.
- d. *-om* is the general verbal marker (e.g., *kopom* 'buy'), although not always used. Verbs exhibit no markers for tense or person. A special *po*+V construction represents a possible TMA (Tense, Mood, Aspect) device.
- e. *-a* tends to mark nouns (e.g., *fiska* 'fish'), which are not inflected and have no plurals.
- f. There is no copula.
- g. The vocabulary derives mostly from Norwegian and Russian, but contains a number of lexical items from other European languages (e.g., *slipom* 'sleep', from English).
- h. RN has SVO syntax. Sentences with adverbial(s) are, however, verb final (e.g., *moja kopom fiska* 'I buy fish'; *moja po tvoja fiska kopom* 'I buy fish from you'). Most sentences are combined paratactically, but embedding is attested. The syntactic possibilities are quite restricted. Most syntactic variation is found in interrogative sentences, RN was used mainly to make inquiries about prices and barter for merchandise.

Bibliography

- Broch I & Jahr E H (1981). *Russenorsk—et pidginspråk i Norge*. Oslo: Novus. (2nd edn. 1984).
- Broch I & Jahr E H (1984). 'Russenorsk: A new look at the Russo-Norwegian pidgin in northern Norway.' In Ureland P S & Clarkson I (eds.) *Scandinavian language contacts*. Cambridge: CUP. 21–65.
- Jahr E H (1996). 'On the pidgin-status of Russenorsk.' In Jahr E H & Broch I (eds.) *Language contact in the Arctic: Northern pidgins and contact languages*. Berlin-New York: Mouton de Gruyter. 107–122.

Jahr E H (2003). 'The emergence of a TMA grammatical device in a stable pidgin: The Russenorsk preverbal *po* construction.' *Journal of Pidgin and Creole Languages* 18, 121–130.

Lunden S S (1995). *The Vardø Merchants' reduced Russian: A. Hansen's Norwegian-Russian vocabulary (1862)*. Oslo: Solum.

Russian

D Ward, University of Edinburgh, Edinburgh, UK

© 2006 Elsevier Ltd. All rights reserved.

The Written Language

Diglossia

The adoption of Eastern Christianity in the 10th century brought to the East Slavs the religious language of the Slavs, Old Church Slavonic (OCS), written in Cyrillic.

Syntax, phraseology, and much of the word formation of OCS owed much to Byzantine Greek. In a russified form, OCS served for centuries as the language of 'culture' of the Russians. The earliest extant text is an aprakos Gospel compiled in 1056–1057 by Deacon Grigorij for Prince Ostromir ('Ostromir Gospel'). Secular works – writs, treaties, codes of law (e.g., *Russkaja Pravda* 'Russian Law', mid-11th century, earliest extant copy 1282), etc. – were written in vernacular Russian.

18th Century

Church Slavonic (CS) and Russian now merged to provide the foundation for the modern literary language. Though the everyday language that V. K. Trediakovskij (1703–1769) advocated as a literary language cannot be found in his own writings, he amply demonstrated the word-forming capabilities of CS elements and processes. M. V. Lomonosov (1711–1765) wrote the first complete grammar of Russian as Russian (1757), distinguishing 'high style' forms (i.e., of CS origin) from the rest and insisting elsewhere that CS words were an ineradicable part of Russian.

Writers of the late 18th and early 19th centuries, e.g., N. M. Karamzin (1766–1826) and others, created a 'new style' (*novyj slog*), in which clarity and straight-forwardness were fundamental criteria, eradicating the ponderous, convoluted earlier 18th century prose style. French provided a model for sentence structure and element order. Karamzin himself produced many new words – straight loans, calques (many based on French) and new creations using the resources of Russian.

Thus, modern literary Russian may be said to be at base a blend of a Graecized Church Slavonic, vernacular Russian, and French syntax and order.

Church Slavonic Features

Almost any printed page of modern Russian reveals numerous elements of CS origin. Such are the nominative singular masculine ending of the adjective *-yj/-ij*, the present active participle in *-ushchij*, etc., (and in general the use of participles), suffixes such as *-ie*, *-stvol-estvo*, *-tel'*, and compound suffixes such as *-eniel-aniel-janie*.

Some morphophonemic alternations show CS origin. For example, CS are *d ~ zhd*, *t ~ shch* against Russian *d ~ zh*, *t ~ ch*. Compare *pobedit'* (PRFV) *~ pobezhdat'* (IMPFV) 'to conquer' with *brodit'* 'to ferment' *~ brozhenie* 'fermentation,' and *obet* 'promise' *~ obeshchat'* 'to promise' with *otvetit'* (PRFV) *~ otvechat'* (IMPFV) 'to answer'. In the striking alternation of pleophonic (*polnoglasnyj*) forms, of Russian origin, and aplophonic (*nepolnoglasnyj*) forms, of CS origin, a vowel *o* or *e* flanks both sides of *l* or *r* in pleophonic forms, whereas a single vowel *a* or *e* follows *l* or *r* in aplophonic forms. Thus, *moloko* 'milk' – *mlekovitaushchij* 'mammalian', *Mlechnyj put'* 'Milky Way'; *korotkij* 'short', *ukorotit'* 'to shorten' – *kratkij* 'brief', *prekratit'* 'to curtail'; *golos* 'voice', *golosovye svjazki* 'vocal cords' – *glasnyj* 'vowel', *soglasnyj* 'consonant'; *bereg* 'bank' – *bregoukre- plenie* 'reinforcement of banks', *bezbrezhnyj* 'boundless'. Pleo- phonic forms are 'concrete,' mundane; aplophonic forms are 'abstract,' 'learned,' 'technical.'

Phonetics

Old Russian had 12 vowel phonemes and some two dozen consonant phonemes, with open syllables and few clusters. The lapse, from the 12th century, of two ultrashort vowels in certain positions initiated the development toward a language with five vowel phonemes, many more consonant phonemes, many clusters and closed syllables, and a system in which palatalization is largely independent of the following vowel, i.e., is largely phonemic.

The vowels /i, e, a, o, u/ have several allophones each, depending on location of stress, consonantal environment, or the two combined. For example, /a/ – *dast* [dast] ‘he will give,’ *dal* [da + l] ‘gave’ MASC, *pjat’* [pæt] ‘five,’ *dala* [dɔ’la] ‘gave’ FEM, *uydat’* [’vidæt] ‘to give out.’

The accent is not fixed and is mobile, shifting in regular patterns in both declension and conjugation, e.g., *storoná* ‘side’, ACC *stóronu*, GEN *storóný*, NOM PL *stórony*, GEN PL *storón*, DAT PL *storónám*, etc.

Except as described below, /o/ is replaced in unstressed syllables by /a/, in a system known as *akan’e* ‘a-saying’ (operating also in southern dialects and Belorussian but not in northern dialects or Ukrainian). Thus ‘town’ appears as *gorod* /’gorat/ NOM SING, *goroda* / gara’da/ NOM PL, *mezhdugorodnyj* /mɨʒduga’rodnij/ ‘interurban’. The last example also illustrates *ikan’e* ‘i-saying’, in which /e/ is replaced in unstressed positions by /i/ (cf. *mezhdú* /’mɛʒdu/ ‘among, between’). *Ikan’e* also affects, in pretonic positions, /a/ after palatalized consonants and /j/, and /o/ after palatalized consonants, /j/ and the palatals /ʃ/ and /ʒ/. Thus, *pjat’* /p at/ ‘five’ ~ *pjati* /pi’ti/ GEN, *let* /lot/ ‘flight’ ~ *letat’* /li’tat/ ‘to fly’, *zheny* /’ʒoni/ ‘wives’ ~ *zhena* /ʒi’na/ ‘wife’. The orthography ignores both *akan’e* and *ikan’e*.

There are 13 pairs of distinctively nonpalatalized/palatalized consonants: /p–pʲ/, /b–bʲ/, /m–mʲ/, /f–fʲ/, /v–vʲ/, /t–tʲ/, /d–dʲ/, /s–sʲ/, /z–zʲ/, /n–nʲ/, /l–lʲ/, /r–rʲ/, /k–kʲ/. Consonants /g/ and /x/ have palatalized allophones; /tʃ/ and /ʃ/ (realized as [ʃʃ]) are nondistinctively palatalized. In addition, there are /ts/, /ʃ/, /ʒ/, and /j/.

Voiced consonants except sonants are devoiced word-finally and before voiceless consonants. Conversely, voiceless consonants are voiced before voiced consonants except sonants, /v/ and /x/. Nonpalatalized consonants are frequently replaced by corresponding palatalized consonants before palatalized consonants, especially homorganic ones. Apart from a very few items and the devoicing of /z/ in prefixes, (e.g., *raz-* ~ *ras-*, *iz-* ~ *is-*), the orthography entirely ignores the various consonant assimilations and final devoicing. Thus, *otdat’* /ad’da’t/ ‘to give back’, *sdelat’* /’zɔ’da’t/ ‘to do’, *gorod* /’gorat/.

Grammar

Nouns

The Old Russian system of eight declensions, three numbers, and seven cases has been simplified into a system of three principal declensions and a vestigial consonant-stem declension, two numbers, and six cases, the dual number and the vocative case having been discarded.

The ‘feminine’ declension in *-al-ja* declines in the singular thus: NOM *komnata* ‘room’, GEN *komnaty*, DAT *komnate*, ACC *komnatu*, INSTR *komnatoj*, PREP (*v*) *komnate*. A few masculine nouns are found in this declension. A typical noun of the ‘masculine’ declension is *stol* ‘table’, *stola*, *stolu*, *stol*, *stolom*, (*na*) *stole*. Neuters decline as masculines except for nominative and accusative, e.g., *okno* ‘window’, PL *okna* (cf. *stoly*) and, usually, GEN PL – cf. *stolov* – *okon*. One masculine noun is still found in the declension that is now otherwise feminine, illustrated by *chast’* ‘part’, *chasti*, *chasti*, *chast’*, *chast’ju*, (o) *chasti*.

Nouns of the masculine declension denoting animate beings use the genitive as an accusative, thus *muzh* ‘husband,’ GEN and ACC *muzha*. The genitive-accusative also applies to all nouns denoting animate beings in the plural, of whatever gender: *zhena* – GEN ACC PL *zhen* (/ʒon/).

Remnants of old declensions include an additional genitive in *-u* of some masculine nouns (usually partitive in function): *kilo sakharu* ‘a kilo of sugar’, cf. *vkus sakhara* ‘taste of sugar’; and an extra PREP case in *-ú* of some masculine nouns, having purely locative function: *v lesu* ‘in the wood’ (cf. *o lese* ‘about the wood’).

A vestige of the dual probably explains the NOM PL MASC in *-á* instead of *-y*, e.g., *goroda*, cf. *stoly*. The graphic identity of the old NOM dual MASC in *-a* with the GEN SING in *-a* of the same declension has led to the use of the GEN SING of a noun of any gender with the numerals *dva* ‘two’, *tri* ‘three’, *chetyre* ‘four’, and higher numerals ending in these elements. Numeral syntax is further complicated by the use of NOM SING with all numerals ending in *odin* ‘one’ and GEN PL with all other numerals: *dva stola* ‘two tables’, *tridsat’ tri stola* ‘33 tables’, *sorok chetyre stola* ‘44 tables’, *sto odin stol* ‘101 tables’, *pjat’ stolov* ‘five tables’.

The genitive not only is obligatory in negative partitive expressions – *Net otveta* (GEN SING) ‘There is no reply’, *Deneg* (GEN PL) *ne khvataet* ‘There isn’t enough money’ – but also is more frequent than the accusative with negated transitive verbs – *Shkoly* (GEN SING) *ona ne brosit/Shkolu* (ACC SING) *ona ne brosit* ‘She will not give up school’.

Syntactically interesting too is the predicative instrumental, standard with certain copula-like verbs and the future of *byt’* ‘to be’: *Ona okazalas’ /stala/ budet sirotoj* ‘She turned out to be /became/ will be an orphan’. With the past tense of *byt’* both the instrumental and the nominative are found: *V to vremja ja byl student(om)* ‘At that time I was a student’, the nominative being more colloquial. *Byt’* has no present tense: *Ona sirota* ‘She (is) an orphan’.

There is no article, definite or indefinite. The 'long' form of the adjective, with a declension different from that of nouns, originally expressed 'definiteness' but is now simply the basic form of the adjective and the only attributive form. The 'short' form no longer declines and is restricted to predicative function, where it simply assigns a property to a subject – *Solntse velika, a Zemlja mala* 'The Sun is big but the Earth is small'. The long form is also used predicatively, assigning the subject to a class of like entities. Compare *Vera očen' umna* (short form) 'Vera is very clever' to *Vera očen' umnaja* (long form) 'Vera is (a) very clever (person)'. This distinction, while still active, is being eroded, especially in colloquial Russian, in favor of the long form.

Verbs

The aspect system of imperfective versus perfective, already active in Old Russian, has led to the reduction of the multiple tenses of Old Russian to just three: past, IMPFV or PRFV, present, IMPFV only, and future, IMPFV or PRFV. The past, originally a periphrastic participial form, is now reduced to what was the participle and so changes according to gender and number, while present and future have 'true' conjugations of three persons and two numbers, the future imperfective being periphrastic. Thus, 'to infringe' – IMPFV *narushat'*, PRFV *narushit'*: past MASC *narushal'/narushil*, FEM *narushala'/narushila*, NEUT *narushalo'/narushilo*, PL *narushali'/narushili*, present *narushaju, narushaesh', narushaet, narushaem, narushaete, narushajut*: FUT IMPFV *budu/budesh'/budet/ budem/ budete/budut narushat'*, FUT PRFV *narushu, narushish', narushit, narushim, narushite, narushat*. The two aspects are differentiated formally by prefixation, suffixal changes, or a combination of the two and occasionally by suppletion. A complication is the existence of many verbs that are not members of minimal pairs, distinguished only by aspect. These form the groups known as *sposoby dejstvija*, Aktionsarten, 'modes of action'. While associated with a base verb, each Aktionsart, appearing in one aspect only, adds a nuance to the base verb, without forming a plain aspectual counterpart. For instance, *stuchat'* 'to knock' is IMPFV and has no plain PRFV counterpart; *postuchat'* PRFV is diminutive or attenuative – 'to knock a little / for a short time' and may have to serve in lieu of a plain PRFV; *stuknut'* PRFV is semelfactive – 'to give a single knock'; *zastuchat'* PRFV is inceptive – 'to start to knock'; *prostuchat'* PRFV is perdurative – 'to knock for a certain period of time'; *postukivat'* IMPFV is intermittent (-diminutive) – 'to knock (a little) from time to time'.

The dozen or so pairs of 'verbs of motion,' while participating in the aspect system, also distinguish between determinate (motion in a single direction) and indeterminate (motion not restricted so): *On letel v Moskvu* 'He was flying to Moscow' – *On letal v Moskvu* 'He flew to Moscow (and back, or several times)'.

There are five participles: PRES ACT *narushajushchij* 'infringing', PRES PASS *narushaemyj* 'being infringed', PAST ACT IMPFV *narushavshij* 'were infringing', PRFV *narushivshij* 'having infringed', and PAST PASS PRFV *narushennyj* 'infringed'. They decline as adjectives and the two passive ones have short forms, the PRFV PASS short form being an indispensable component of the passive voice: *Zakon byl narushen*. 'The law was infringed'. The two indeclinable adverbial participles, often called gerunds, are, for example, IMPFV *narushaja* 'infringing' and PRFV *narushiv (shí)* 'having infringed'. Subordination by means of participles and gerunds, instead of relative and adverbial clauses, is common.

Lexis

While the bulk of the lexis is Slavonic, Russian has not been averse to borrowing at all periods. From Western European languages Dutch has provided nautical terminology: *botsman* 'bosun', *kil'vater* 'wake'; German – military and other terminology: *lager* 'camp', *landshaft* 'landscape', *butterbrod* 'sandwich'; French – military, mundane and cultural vocabulary: *batal'on* 'batallion', *pal'to* 'overcoat', *rezhisser* 'producer'; English – nautical terms: *midshipman* 'midshipman', mundane: *bifshteks* 'steak', industrial: *rel'sy* 'railway lines', sociopolitical: *bojkot* 'boycott', *khuligan* 'hooligan', and in the 20th century, sport: *futbol* 'football', *vindsorfiging* 'windsurfing', and technical: *bul'dozer* 'bulldozer', *komp'juter* 'computer'.

Naturally, Russian has gone on exploiting its historically established word-forming processes, but it has also exploited less traditional ones. In this respect, notable are appositional compounds such as *raketa-nositel'* 'carrier rocket', *dom-muzej* 'home (which is also a) museum', and above all acronyms and various other accreted abbreviations – *vuz* (*vysshee uchebnoe zavedenie*) 'higher educational institution', *GUM* (*Gosudarstvennyj Universal'nyj Magazin*) 'State Department Store', *ROSTA* (*Rossijskoe Telegrafnoe Agentstvo*) 'Russian Telegraph Agency', *kolkhoz* (*kollektivnoe khozajstvo*) 'collective farm', *univermag* (*universal'nyj magazin*) 'department store', *zarplata* (*zarabotnaja plata*) 'wages', *fizkul'tura* (*fizicheskaja kul'tura*) 'physical training'.

Influence of Russian

In varying degrees, Russian has provided loanwords, especially relating to 20th century life, of technological and cultural significance for many non-Slavonic languages of the former Soviet Union. An extreme case of such borrowing from Russian is provided by Chukchi. In Altaic, North Caucasian, and easterly Uralic languages, subordinating constructions on the Russian model have become common. The languages of many small speech-communities (Ingrian, Veps, Vot, Mordvinian, Siberian languages, etc.) have retreated or are retreating in the face of Russian.

Bibliography

- Comrie B & Stone G (1978). *The Russian language since the Revolution*. Oxford: Oxford University Press.
- Isacenko A V (1962). *Die russische Sprachen der Gegenwart, Teil I – Formenlehre*. Halle: Niemeyer.
- Issatschenko A (1980–1983). *Geschichte der russischen Sprache*. Heidelberg: Carl Winter.
- Unbegaun B (1960). *Russian grammar*. Oxford: Oxford University Press.
- Ward D (1981). 'Loan-words in Russian.' *Journal of Russian Studies* 41, 3–14, 42, 5–14.

Ryukyuan

M Shibatani, Rice University, Houston, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

Also known as Luchuan and Okinawan, the Ryukyuan (Okinawan, Central) language comprises a group of diverse dialects of the former Ryukyu Kingdom, 1429–1879, which has lost much of its political and economic independence since 1609, when it fell to the hands of the Shimazu clan of Kagoshima, Kyushu. Following the Japanese annexation in 1879, the Ryukyu Islands became a prefecture of Japan – *Okinawa ken* 'Okinawa Pref.' – the status it regained in 1972, when the islands were returned to Japan from the American occupation. The Japanese government policy of fostering the use of standard Japanese since the time of the Meiji restoration (1868) has helped marginalize local dialects throughout Japan, and it has also had a pronounced effect in Okinawa. Based on the most recent census (2002), it can be estimated that of the current population of 1.3 million Okinawans, less than 300 000 people over the age 50 speak some variety of the Ryukyuan language with varying degrees of proficiency. Since children no longer learn to speak Ryukyuan, the language is bound to become extinct within the next 50 years unless active revitalization efforts are mounted.

Hypothesizing the sister-language relationship, Chamberlain (1895) remarked that the relationship between Ryukyuan and Japanese is something like that between Spanish and Italian or that between French and Italian. But unlike these Romance languages, the Ryukyuan dialects are often mutually completely unintelligible among their speakers, let alone to the speakers of any mainland dialect. Japanese dialectologists,

on the other hand, have generally regarded Ryukyuan as a branch of Japanese dialects comprising three large groups: the Amami-Okinawa group (Amami dialect, Okinawan dialect), the Miyako-Yaeyama group (Miyako dialect, Yaeyama dialect), and the southernmost Yonaguni dialect group. The dialect of Shuri, the former capital of the Ryukyuan Kingdom, of the main Okinawa Island is generally regarded as the standard Ryukyuan, and it has served as a lingua franca of the Ryukyus.

It is generally estimated that the Ryukyuan stock split from the mainstream Japanese language at the latest around the 6th century A.D. (Hattori, 1976). Ryukyuan dialects show systematic sound correspondences with the modern Tokyo dialect, and they preserve a number of distinct features of Old Japanese. As shown in Tokyo–Shuri correspondences such as *ame:ami* 'rain', *hone:funi* 'bone', and *kokoro:kukuru* 'heart', the mid vowels have been raised in Ryukyuan dialects with the result of the standard five vowels *i, e, u, o, a* being reduced to the three vowels *i, a, u*. Innovative phonological developments include palatalization of /k/ before the /i/ corresponding to the Tokyo /i/ (Shuri *tfiri* : Tokyo *kiri* 'fog'), centralization of the original /i/ in certain dialects (e.g., Miyako, Yaeyama), and the development of long mid vowels /o:/ and /e:/ from *au*, *ao*, and *ou*, as well as *e*: from *ai* and *ae*, respectively.

The features of Old Japanese preserved in Ryukyuan dialects cover all aspects of grammar. The Old Japanese consonant /p/ is preserved in such words as *piru* 'day time', *pi*: 'fire', and *pa*: 'leaf', corresponding to the Tokyo forms *hiru*, *hi*, and *ha*, respectively. The Ryukyuan lexicon contains older forms such as *tudzi* 'wife', *wan* 'I', and *warabi* 'child'. The notable

syntactic features of Old Japanese preserved include the distinction between the conclusive form and the attributive form of verbs and adjectives; e.g., *katfun* 'write-Conclusive' and *katsuru* 'write-Attributive' correspond to the Tokyo form *kaku* used in both conclusive and attributive functions. Also seen is the preservation of the nominative function of the particle *nu* (Old Japanese/Modern Japanese *no*) in the main clause. In the total picture of the Japanese dialects, the Ryukyuan dialects form the most peripheral groups that preserve historically residual forms in line with the classical theory dialectology.

Bibliography

- Chamberlain B (1895). *Essay in aid of a grammar and dictionary of the Luchuan language*. Transaction of the Asiatic Society of Japan 23 (suppl.).
- Hattori S (1967). 'Ryuukyuu hoogen to hondo hoogen.' In *Okinawagaku no Reimei*. Tokyo: Okinawa Bunka Kyokai. 7–55.
- Kokuritsu K K (1963). *Okinawago Jiten*. Tokyo: Okura-sho.
- Shibatani M (1990). *The languages of Japan*. Cambridge: Cambridge University Press.
- Tojo M (1927). *Wagakuni no hoogen-kukaku*. Tokyo: Ikueishoin.

This page intentionally left blank

S

Saami

P Sammallahhti, University of Oulu, Oulu, Finland

© 2006 Elsevier Ltd. All rights reserved.

Saami is a subfamily of closely related languages within the Uralic phylum. At present, the Saami languages are spoken in an area arching from Dalecarlia in central Sweden to the tip of the Kola peninsula in northwestern Russia. The number of native speakers is ca. 30 000; ca. 85% of these speak North Saami (Saami, Northern) in Finland, Norway, and Sweden. The rest are unevenly distributed between Lule Saami (Saami, Lule) (south of North Saami in Norway and Sweden, estimated 2000 speakers), Kildin Saami (Saami, Kildin) (inland and northern coast of the Kola peninsula, 900 speakers), South Saami (Saami, Southern) (in the southernmost Saami areas in Norway and Sweden, 500 speakers), Skolt Saami (Saami, Skolt) (in Finland and some speakers in Russia, 400 speakers), Inari Saami (Saami, Inari) (in Finland, 300 speakers). The rest, Ume Saami (Saami, Ume) and Pite Saami (Saami, Pite) in Sweden between South Saami and Lule Saami areas, and Ter Saami (Saami, Ter) in eastern Kola peninsula are maintained by a score of old speakers each; Akkala Saami (Saami, Akkala) in southwestern Kola peninsula is probably extinct; the documentation of the minor languages is unsatisfactory (Ter, Pite, Ume) or highly unsatisfactory (Akkala).

There have been Saami idioms down to southern Finland and southern Karelia in Russia; of these idioms, the northernmost ones went into oblivion in the 19th century. There is evidence (language documents, etc.) of Saami presence south of the South Saami area in Sweden and Norway. Loanwords from Finnic, Proto-Indo-European, Aryan, Germanic, Baltic, and Slavic witness contacts with other linguistic groups. There is an ostensible non-Uralic substrate, especially in place names.

First Saami books date to 1619. The author belonged to a trader family and the language represents a pidgin with Saami, Finnish, and Swedish words and hardly any inflection. The first books representing Saami vernaculars (Lule Saami, Ume

Saami) were published in the same century. At present, six Saami languages (South, Lule, North, Inari, Skolt, and Kildin) have a literary norm.

The Saami languages are structurally close to the rest of the Uralic languages. The finite verb agrees with the subject and is conjugated in three numbers (singular, dual, and plural), three persons (first, second, and third), and two tenses (the present/future and the preterit) in most languages; the auxiliary *leat* 'to be' together with nonfinite forms of the main verb is used to form aspectual compound forms (progressive vs. terminative aspect in both tenses, e.g., *lean boahtime* 'I am coming' vs. *lean boahtán* 'I have come'). The number of morphological moods varies from two (indicative, imperative) to five (indicative, conditional, potential, imperative, adhortative).

Negation is expressed by a negative verb; in the idioms southwest of North Saami, it has two tenses. In North Saami and east of the language area, tense is encoded in the nonfinite main verb (e.g., North Saami *in mana* 'I do not go' vs. *in mannan* 'I did not go'). The rest of the auxiliaries (mostly for epistemic and deontic modalities) show a more complete conjugation. In addition to compound tenses the nonfinite verb forms are also used for sentence-embedding, e.g., *Máret logai Máhte* [AccSg] *boahtán* [Perfect Participle] 'Mary said that Matthew has come.' Derivation is extensive within and across word classes; in verbs, there are several morphological passives and causatives in addition to a wide selection of aspectual derivatives (frequentative, subitive, etc.).

The Saami languages are nominative-accusative languages; in North Saami, some verbs denoting natural processes may take their single participant argument either in the nominative or in the accusative (e.g., *biegga* [NomSg] *garai* 'the wind became harder' ~ *garai biekká* [AccSg] id.). The basic word order in most Saami idioms is SVO but the older SOV is still dominant in South Saami, and the object neutrally precedes its nonfinite head (S Aux O V). Word order is free for the dependents of the verb and determined by pragmatic factors; the attribute precedes its head; postpositions dominate over prepositions.

In addition to the grammatical cases nominative (= no case) and accusative, nominals have the local cases illative, inessive, and elative (illative and locative in North Saami and the languages east of its area), the predicative cases comitative, essive, and abessive; as a rare case of 'degrammaticalization,' the abessive ending has evolved into a postposition in North Saami (e.g., **guolihaga* > *guoli haga* 'without fish'). The local cases are also used in nonlocal arguments (e.g., *Máhhte ballá gumpes* [LocSg] 'Matthew is afraid of the wolf'); the comitative also expresses the instrument argument. In addition to determinative (e.g., *Máhhte* [GenSg] *govva* 'Matthew's picture') and complementing uses (e.g., *seainni* [GenSg] *čáda* 'through the wall'), the genitive expresses the owner argument when the theme is definite, e.g., *beana* [NomSg] *lea áhči* [GenSg] 'the dog is father's'; if the theme is nondefinite, the locative (< inessive) is the case of the owner argument, e.g., *áhčis* [LocSg] *lea beana* [NomSg] 'father has a dog'; in South Saami, genitive is used in both cases *aehtjien* [GenSg] *bienje* [NomSg] 'father has a dog,' *bienje aehtjien* 'the dog is father's.'

Saami phonology is an extreme sport: a bisyllabic stem may have over 20 different phonological forms depending on grade alternation, compensatory lengthening, vowel balance and metaphony, etc., caused by different suffixes. The number of consonant phonemes is 19–40 depending on idiom, and the basic vowel phonemes (5–10 depending on idiom) are combined to form vowel sequences (5–10 geminate vowels and 4–10 diphthongs with the first component higher than the second, e.g., /ie/ and /oa/). Word stress

is trochaic, e.g., /kú.laa.hàl.laa.pèeh.teh/ 'you understand each other,' but morphology may cause deviations from the rule, e.g., /múj.j.hta.liš.kòah.tiih/ 'to begin telling' (/š.kòah.tii-/ is a derivational ending). Ume, Pite, Lule, North, Inari, and Ter Saami have three degrees of quantity in consonants, e.g., North Saami /čáal.l.iih/ 'writers' (with an extra syllabic pulse) vs. /čáal.liih/ 'to write' vs. /čáa.liih/ 'make him/her write!' Also vowels in stressed syllables show three contrasting lengths (roughly [a] – [a:] – [a:~]) but these derive from the phonological oppositions (a) single vowel vs. vowel sequence and (b) initial vs. final stress in a vowel sequence, e.g., /sáh.te/ 'haphazard' – /ij/ maáh.te/ 'does (not) know how to' – /máah.te/ 'Matthew's (GenSg)' (phonetically [sahte] – [ma:hte] – [ma:hte]); initial vs. final stress is also found in diphthongs, e.g., /sóa.đan/ 'I fight' vs. /poá.đan/ 'I come.' These contrasts originate in the grammaticalization of allegro forms in which vowel sequences in stressed syllables receive final stress, and vowel sequences in the following syllables are reduced to single vowels (* /póa.đaan/ > /poá.đan/ 'I come'). Syllable border placement is distinctive in at least North Saami, e.g., eastern North Saami /pól.htuuh/ 'to rummage' vs. /pólh.tuuh/ 'you rummage.'

Bibliography

- Sammallahti P (1998a). 'Saamic.' In Abondolo D (ed.) *The Uralic languages*. London/New York: Routledge. 43–95.
Sammallahti P (1998b). *The Saami languages: an introduction*. Kárášjohka/Vaasa: Davvi Girji.

Salishan Languages

S Thomason, University of Michigan, Ann Arbor, MI, USA

© 2006 Elsevier Ltd. All rights reserved.

The 23 languages of the Salishan family are spoken in the U.S. Northwest and neighboring British Columbia, along the Pacific coast in Washington, Oregon, and British Columbia, and inland to interior British Columbia, the Idaho panhandle, and northwestern Montana. The languages fall into five distinct branches, according to the most commonly accepted subgrouping schema (Czaykowska-Higgins and Kinkade, 1998a; 3): **Bella Coola**, the northernmost language and a one-language branch; **Central** (or

Coast) **Salish**, comprising Comox–Sliammon (Comox), Pentlatch, Sechelt, Squamish, Halkomelem, Northern Straits, Klallam (Clallam), Nooksack, Lushootseed, and Twana; the **Tsamosan** languages, Quinault, Lower Chehalis, Upper Chehalis, and Cowlitz, located primarily south of the Central Salish languages; **Tillamook**, a one-language branch, spoken in Oregon; and **Interior Salish**, which is divided into two branches: the three Northern Interior languages – Shuswap, Lillooet, and Thompson River Salish – are spoken in interior British Columbia, and the four Southern Interior languages, Colville–Okanagan, Columbian, Coeur d'Alene, and Montana Salish (a.k.a. Flathead-) Kalispel–Spokane (Kalispel = Pend d'Oreille), are spoken primarily in the interior U.S.

Northwest (although Colville–Okanagan is also spoken across the border in Canada). In several instances, as with Montana Salish–Kalispel–Spokane, different tribes speak closely related dialects of a single (nameless) language.

Various proposals have linked the Salishan family genetically to other Northwest languages, but none of these is widely accepted. The isolate Kutenai, which has long been in close contact with some of the Southern Interior languages, is one candidate for a distant relative. Other proposed congeners are the Wakashan and Chemakuan families, also located in the Pacific Northwest; together, Salishan, Wakashan, and Chemakuan comprise the core of the famous Pacific Northwest linguistic area. A number of striking typological features are found in all three of these families (and some of them also in Kutenai); most of the features mentioned below for Salishan also occur in the other two core Pacific Northwest families.

All Salishan languages have rich consonantal inventories that include ejectives, lateral obstruents, velar vs. uvular obstruents, labialized dorsal consonants, and (in some of the languages) glottalized resonants and pharyngeal consonants. **Table 1** shows a widely (though not universally) accepted set of Proto–Salishan consonant phonemes (modified from Kroeber, 1999: 7, partly on the basis of comments in Czaykowska-Higgins and Kinkade, 1998a: 51–52). Vowel inventories, in sharp contrast, are relatively simple: Proto–Salishan is generally believed to have had just four vowel phonemes, /i ə a u/.

Salishan phonology displays other striking features as well, notably the presence (in almost all the languages) of very elaborate consonant clusters, as in Montana Salish *Ta qesm'l'm'él'čstmstx^w!* ‘Don’t play with that!’ Another widely shared phonological phenomenon is a sound change, in most of the languages and apparently independently in at least two subgroups, from the velar consonants *k k' x* to alveopalatals *č č' š* (and then sometimes to other consonants later).

Morphologically, Salishan is heavily agglutinative, or polysynthetic. All the languages have many suffixes, including both grammatical suffixes – for example, transitivizers, subject markers, and object makers – and lexical suffixes by the dozens, primarily

indicating concrete objects (e.g., ‘hand,’ ‘face/fire,’ ‘nose/road/cost,’ ‘round object,’ ‘root/berry’). Prefixes are not so numerous, though most of the languages have locative prefixes and several others as well. An affix-loaded Montana Salish word, for instance, is *q^wo č-tač^wl-m-nt-cút-m-nt-m* ‘they came up to me’ (lit. ‘me to-START-derived.transitive-transitive-reflexive-derived.transitive-transitive-indefinite.agent’). This word contains one locative prefix and six suffixes, with one suffix set, *-m* ‘derived transitive’ plus *-nt* ‘transitive,’ repeated after the reflexive suffix etransitivizes the stem.

Reduplication is a prominent morphological process in Salishan, used for such purposes as distributive plural (e.g., Montana Salish *qe č'uč'úw* ‘all of us are gone, we left one at a time’ vs. *qe č'úw* ‘all of us are gone, we left in a group’) and diminutive. Salishan languages have pronominal clitics that mark certain subjects (e.g., in intransitive predicates), suffixes that mark other subjects (e.g., in transitive predicates), suffixes that mark patients, and pronominal possessive affixes (see Czaykowska-Higgins and Kinkade, 1998a: 31).

Word classes include at least full words and particles. Because every full word can serve as the predicate of a sentence, some scholars have argued for the absence of a lexical distinction between verbs and nouns (see Kuipers, 1968; Kinkade, 1983; Jelinek, 1998; for the other side of this controversy, see Van Eijk and Hess, 1986; Kroeber, 1999: 33–36). There is general agreement that, if the distinction exists, its morphological and syntactic ramifications are weaker than in most or possibly all language families outside the Pacific Northwest. Salishan languages have suppletive lexical pairs of roots with singular and plural reference, e.g., Montana Salish *čn řútč^w* ‘I went in’ vs. *qeř npílš* ‘we went in.’

Nearly all Salishan languages are predominantly predicate-initial, mostly VSO but in some languages VOS; word order is rather free. In all the languages transitivity is a major morphosyntactic category, with transitivizing and detransitivizing suffixes, applicatives, causatives, and other complexities; they are head-marking. Jelinek (e.g., 1984) and Jelinek and Demers (e.g., 1994) have proposed that these are pronominal argument languages, with full noun phrases having the status of adjuncts rather than arguments. This claim has been debated vigorously, on both sides of the issue, by Salishanists and other theoreticians.

Research on Salishan languages began early, with wordlists collected by travelers as early as 1810 and the first grammar and dictionary published later in the 19th century – a grammar and a thousand-page two-volume dictionary of Montana Salish (Mengarini,

Table 1 Proto–Salishan Consonant Phonemes

| | | | | | | | | |
|----|------|----|----|------|----------------|----|----------------|----------------|
| p | t | | c | k | k ^w | q | q ^w | ʔ |
| p' | t' | χ' | c' | k' | k ^w | q' | q ^w | |
| | | ʈ | s | x | x ^w | ç | ç ^w | (h) |
| m | n | | | | | | | |
| | (r) | l | y | (ʏ) | w | | ʃ | ʃ ^w |
| | (r') | l' | y' | (ʏ') | w' | | ʃ' | ʃ ^w |

1961; Mengarini *et al.*, 1877–1879). Modern Salishan linguistics has been flourishing for over half a century, and three especially important surveys have appeared: Thompson, 1979; Czaykowska-Higgins and Kinkade, 1998a and Kroeber, 1999 (in particular Chap. 1). An annual conference, Salish and Neighboring Languages, is held each August, and the conference preprints are a major source for information on the languages. A sizable number of descriptive grammars and dictionaries of various Salishan languages are now available, together with a great many more articles on specific theoretical and descriptive issues, a large monograph on comparative syntax (Kroeber, 1999), and an etymological dictionary (Kuipers, 2002).

All the Salishan languages are gravely endangered. Czaykowska-Higgins and Kinkade (1998a: 64–67) report speaker figures that range from about 500 (for 4 languages) to fewer than 10 (for 9 languages) and 0 (for several now-vanished languages). Language-revitalization efforts are under way, however, for many of the Salishan languages.

Bibliography

- Czaykowska-Higgins E & Kinkade M D (1998a). Salish language and linguistics. In Czaykowska-Higgins & Kinkade (eds.), 1–68.
- Czaykowska-Higgins E & Kinkade M D (eds.) (1998b). *Salish languages and linguistics: theoretical and descriptive perspectives*. Berlin: Mouton de Gruyter.
- Jelinek E (1984). Empty categories, case, and configurationality. *Natural Language and Linguistic Theory* 2, 39–76.
- Jelinek E (1998). ‘Prepositions in Northern Straits Salish and the noun/verb question.’ In Czaykowska-Higgins & Kinkade (eds.), 325–346.
- Jelinek E & Demers R A (1994). ‘Predicates and pronominal arguments in Straits Salish.’ *Language* 70, 697–736.
- Kinkade M D (1983). ‘Salish evidence against the universality of “noun” and “verb.”’ *Lingua* 60, 25–39.
- Kroeber P D (1999). *The Salish language family: reconstructing syntax*. Lincoln: University of Nebraska Press.
- Kuipers A H (1968). ‘The categories verb–noun and transitive–intransitive in English and Squamish.’ *Lingua* 21, 610–626.
- Kuipers A H (2002). *Salish etymological dictionary*. University of Montana Occasional Papers in Linguistics 16. Missoula: UMOPL Linguistics Laboratory.
- Mengarini G S J (1861). *Grammatical linguae Selicae*. New York: Cramoisy Press.
- Mengarini G S J Giorda J *et al.* (1877–1879). *A dictionary of the Kalispel or Flat-head Indian language*. St. Ignatius, MT: St. Ignatius Print. [This work is usually referred to as Giorda 1877–1879.]
- Thompson L C (1979). ‘Salishan and the Northwest.’ In Lyle Campbell & Marianne Mithun (eds.) *The languages of Native America: historical and comparative assessment*. Austin: University of Texas Press. 692–765.
- Van Eijk J P & Hess T M (1986). Noun and verb in Salish. *Lingua* 69, 319–331.

Samar-Leyte

J W Lobel, University of Hawai‘i, Manoa, Honolulu, HI, USA

© 2006 Elsevier Ltd. All rights reserved.

Waray-Waray (Samar-Leyte) is an Austronesian language of the Central Bisayan subgroup of Central Philippine languages. With approximately 3 million speakers, Waray-Waray ranks sixth in terms of number of speakers in the Philippines, fifth in the Central Philippine subgroup, and third in the Visayan Islands. Waray-Waray is spoken in an area that roughly corresponds to the borders of the Eastern Visayas (Region VIII), except for the western coast of Leyte and Biliran and a number of small islands off the north-western coast of Samar, most of which are Cebuano-speaking, except for Capul Island, which is home to Sama Abaknon, a language of the Sama-Abaknon subgroup.

As a Central Bisayan language, Waray-Waray is most closely related to Ilonggo (Hiligaynon), Capiznon, Masbatenyo, Romblomanon, Central and Southern Sorsoganon, Poroanon, and Bantayanon. Outside of the Central Bisayan subgroup, these languages are related to Cebuano, Asi (Bantoanon), the Western Bisayan languages (including Aklanon, Kinaray-a, and Unhan [Inonhan]), and the Southern Bisayan languages (including Tausug, Surigaonon, and Butuanon). Within the Central Philippine subgroup, the Bisayan languages are coordinate with Tagalog and the Bikol (Bicolano, Central) languages.

The ‘standard’ dialect of Waray-Waray is that of Tacloban City. The Waray-Waray-speaking areas exhibit substantial dialect variation, and in many places no two towns speak the same dialect. Approximately two dozen dialects and subdialects are found in the region. The greatest major dividing line is between Northern Samarenyo and the rest of the

Waray-Waray area. The dialect of Allen, Samar, is predominantly Southern Sorsoganon mixed with Northern Samareño, and neighboring towns also have a considerable amount of borrowing from Southern Sorsoganon. There is also a modest amount of evidence for a split between Samar Waray-Waray and Leyte Waray-Waray, although much of this split consists of borrowings from Cebuano into Leyte Waray. The dialect of Abuyog, Leyte, is particularly heavily influenced by Cebuano, as is the dialect of Culaba, Biliran. It is also interesting that the dialects of the oldest settlements in Baybay, Leyte, (C. Rubino, personal communication), and the Camotes Islands (Wolff, 1967) show a Warayan substratum, indicating that Waray-Waray was much more widespread in previous centuries before the expansion of the Cebuanos in the mid-1800s (Larkin, 1982). In total, there are approximately 25 dialects and subdialects of Waray-Waray, defined mostly by lexical and morphological variation, as very little phonological and grammatical differences exist.

The earliest written works on the Waray-Waray language are Domingo Ezguerra's 1663 grammar *Arte de la Lengua Bisaya de la Provincia de Leite* and a dictionary by Mateo Sanchez (1562–1618) published a century after his death as the *Vocabulario de la Lengua Bisaya* (1711).

Recent works include two dictionaries (Abuyen, 1994; Tramp, 1995), a series of pedagogical texts (Wolff and Wolff, 1967), and two compilations of literature (Luangco, 1982b; Sugbo, 1995). Zorc (1977) contains data from three Waray-Waray dialects in comparison to other Bisayan languages. Waray-Waray is the language of the church throughout the Eastern Visayas region, and by far the most readily available literature in Waray-Waray is religious in nature, including two modern Bible translations and numerous prayer pamphlets.

Waray-Waray has the basic Central Philippine-type phonology, with 16 consonants /p b m w t d n s l r y k g sng₂ uʔ₅ h/ and three vowels /a i u/, and both stress and length are contrastive. Most of the dialects of northeastern and eastern Samar have a fourth vowel as a reflex of PAN *e, a high, central, tense unrounded vowel i. The Waray-Waray orthography is mostly regular except that it does not represent stress, length, or the glottal stop.

Waray-Waray is most readily distinguishable from other Central Philippine languages by the *s > /h/ sound change that has affected a small number of common grammatical morphemes in all areas of Samar south of the Sta. Margarita-Matuginao-Las Navas-Gamay line and all of Leyte Waray except the towns of Javier and Abuyog. However, the *s > /h/ change and the loss or retention of PAN *e are areal features, and therefore do not define genetic subgroups within Waray-Waray.

Waray-Waray is agglutinative, with a complex system of verbal morphology expressing a wide variety of semantic and syntactic contrasts. Although sometimes analyzed as ergative, these languages are probably of a separate type called Symmetrical Voice Languages, in which multiple voice distinctions exist, yet none can be considered more 'basic' than the other (Himmelman, to appear). Like most other Philippine languages, there are four main verbal 'focuses' (actor, object, location, and beneficiary; see Table 1) and three 'case' distinctions (nominative, genitive, and oblique) in noun phrases, name phrases, and pronouns (marked by an introductory morpheme) (see Table 2). Nouns, adjectives, and verbs distinguish between singular, plural, and in some cases, dual, and verbs may also be marked for reciprocal action. A number of other meanings can be marked by verbal affixes, including accidental, abilitative, distributive, causative, social, and diminutive. Tense-aspect-mood

Table 1 Waray-Waray focus-mood-aspect morphology

| | | Actor focus | Object focus (1) | Object focus (2)/beneficiary focus | Location focus |
|------------------|-------------------------------|-------------|------------------|------------------------------------|----------------|
| -um-verbs | Infinitive | -um- | -on | i- | -an |
| | Past/perfective | -inm- | -in- | i-...-in- | -in-...-an |
| | Present/progressive | nā- | -in-R- | i-...-in-R-... | -in-R-...-an |
| | Future | mā- | R-...-on | i-R- | R-...-an |
| | Imperative/subjunctive | ø- | -a | -an | -i |
| mag-verbs | Future subjunctive | R- | R-...-a | R-...-an | R-...-i |
| | Infinitive | mag- | pag-...-on | ig- | pag-...-an |
| | Past/perfective | nag- | gin- | igin- | gin-...-an |
| | Present/progressive | nag-R- | gin-R- | igin-R- | gin-R-...-an |
| | Future | mag-R- | pag-R-...-on | ig-R- | pag-R-...-an |
| | Imperative/subjunctive | pag- | pag-...-i | pag-...-an | pag-...-i |
| | Future subjunctive | pag-R- | pag-R-...-i | pag-R-...-an | pag-R-...-i |

Table 2 Waray-Waray pronouns

| | | Nominative | Genitive | Oblique |
|----------|---------------|------------|---------------------|----------------|
| Singular | 1st | akó, ak | ko | (ha/sa) ákon |
| | 2nd | ikáw, ka | mo, nimo, nim | (ha/sa) imo/im |
| | 3rd | hiyá/siyá | niyá | (ha/sa) iya |
| Plural | 1st exclusive | kamí | námon | (ha/sa) ámon |
| | 1st inclusive | kitá | náton | (ha/sa) áton |
| | 2nd | kamó | niyo | (ha/sa) iyo |
| | 3rd | hirá/sirá | níra | (ha/sa) ira |
| | | | | |

Table 3 Waray-Waray case markers

| | | Standard Waray | Abuyog | Calbayog; Northern Samar-A | Northern Samar-B |
|-----|-------------|----------------|--------|----------------------------|------------------|
| Nom | -ref | in | in | in | i(n) |
| | +ref, +past | an | an | an | a(n) |
| | +ref, -past | it | it | | |
| Gen | -ref | hin | sin | sin | si(n) |
| | +ref, +past | han | san | san | sa(n) |
| | +ref, -past | hit | sit | | |
| Obl | | sa | sa | sa | sa |

Table 4 Waray-Waray demonstratives

| | Nominative | Genitive | Oblique |
|----------------------------------|------------|-----------------|------------------|
| Near speaker; far from addressee | adí | hadi/ sadi | didi, ngadi |
| Near speaker and addressee | ini | hini/sini | dinhi, nganhi |
| Far from speaker; near addressee | iton | hiton/ siton | dida, ngada |
| Far from speaker and addressee | adto | hadto/ sadto | didto, ngadto |

distinctions include infinitive, past/perfective, present/progressive, future, imperative/subjunctive, and future subjunctive. Both reduplication and repetition are productive mechanisms that can denote diminutive, repetitive, and intensive meanings, among others.

Waray-Waray has much the same grammatical structure as Tagalog, except for (a) the existence of distinct imperative forms, (b) a preference for inflecting verbs for plural actors, (c) a more elaborate system of case markers that distinguish between referential and non-referential and past and non-past in both the nominative and genitive cases (see Table 3), and (d) a four-way distinction in demonstratives, including a contrast

between referents that are near only the speaker vs. those that are near both the speaker and the addressee (see Table 4).

A noteworthy feature of the lexicon of dialects of Waray-Waray spoken in parts of eastern and northeastern Samar is the existence of a register of vocabulary reserved for usage by speakers when they are angry.

Bibliography

- Abuyen T (1994). *Diksyunaryo Waray-Waray (Visaya) – English – Tagalog*. Philippines: Tomas Abuyen.
- Alcázar Fr A V (1914). *Diccionario Bisaya-Español para las provincias de Sámar y Leyte compuesto por el M. R. P. Fr. Antonio Sánchez de la Rosa, corregido y aumentado, tercera edición*. Manila: Imp. Y Lit. de Santos y Bernal.
- Cinco E P (1977). 'A contrastive analysis of Waray and Pilipino (Tagalog) adverbials.' M.A. Thesis. Manila: Philippine Normal College.
- Diller T C (1971). 'Case grammar and its application to Waray, a Philippine language.' Ph.D. diss. Los Angeles: University of California.
- Ezguerra P D (1747). *Arte de la Lengua Bisaya de la Provincia de Leite*. Manila: Compañía de Jesus.
- Figueroa A (1872). *Arte del Idioma Visaya de Samar y Leite* (2nd edn.). Binondo, Philippines: Imprenta de Bruno Gonzales Moras.
- Himmelman N (to appear). 'Typological characteristics.' In Himmelman N & Adelaar K A (eds.) *The Austronesian languages of Asia and Madagascar*. London: Curzon Press.
- Larkin J A (1982). 'Philippine history reconsidered: a socio-economic perspective.' *The American Historical Review* 87(3), 595–628.
- Luangco G C (ed.) (1982a). *Kandabao: essays of Waray language, literature and culture*. Tacloban City: Divine Word University Publications.
- Luangco G C (ed.) (1982b). *Waray literature: An anthology of Leyte-Samar writings*. Tacloban City: Divine Word University Publications.
- Macariola G P (1977). 'A contrastive analysis of Waray and Tagalog (Pilipino) major transitive verbs.' M.A. Thesis. Manila: Philippine Normal College.
- Paquio M N (1969). 'An investigation of Waray phonology.' M.A. Thesis. Columbia: University of Missouri.
- Piczon P B (1974). 'Elements of modern Waray.' M.A. Thesis. Manila: Ateneo de Manila University.
- Sanchez M (1711). *Vocabulario de la lengua Bisaya (Samar-Leyte)*.
- Sánchez de la Rosa A (1887). *Diccionario Bisaya-Español para las provincias de Sámar y Leyte*.
- Soriano F (1959). 'Fundamental grammar notes in the Waray Waray dialect.' As presented at the Sanz School of Languages. Cornell University.
- Sugbo V N (1995). *Tinipigan: an anthology of Waray literature*. Manila: National Commission for Culture and the Arts.

- Tramp G D Jr (1995). *Waray dictionary*. Kensington, MD: Dunwoody Press.
- Wolff J U (1967). 'History of the dialect of the Camotes Islands, Philippines, and the spread of Cebuano Bisayan.' *Oceanic Linguistics* 6(2), 63–79.

- Wolff J U & Wolff I O (1967). *Beginning Waray-Waray*. Ithaca: Cornell University.
- Zorc R D (1977). *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Pacific Linguistics C-44. Canberra: Australian National University.

Sango

W J Samarin, University of Toronto, Toronto, Canada

© 2006 Elsevier Ltd. All rights reserved.

Sango was declared the 'national language' in the constitution of the Central African Republic (1964), French alone having the status of 'official language.' Sango also was given this status in 1991, allowing it to be used in governmental communications. In practice, however, it is not yet a legal language and is not used in public education, French being the official medium. Missionaries introduced written Sango in the 1920s, Catholics and Protestants using different orthographies; an official one was established by presidential decree in 1984. Literacy in Sango has been used nonreligiously, primarily in personal correspondence and by radio 'journalists' preparing notes in *ad hoc* orthographies based on religious Sango. There were about 46 hours of broadcasts in Sango in 1994, but most of them were broadcasts of dance music in the Kinshasa style, the rest news and practical information.

Sango is a pidgin in origin, emerging very quickly after representatives of King Leopold II in 1887 arrived to claim land and trade for elephant tusks in the Ubangi river basin, followed in 1889 by the French. Unlike other pidgins, Sango appears to have arisen, not from the attempts of whites to communicate in an indigenous language, but from the attempts of the linguistically diverse African soldiers and workers who were brought to the region for its colonization. The Belgians used many men from the east coast, the French from the west coast, and both many more from the Bantu population along the Congo river. Its existence as a *lingua franca* was noted by Belgians in 1895. It is based on the Ngbandi sub-family of languages (not just the variety called Sango) that make up the larger Ubangian family, to which most of the Central African languages belong. Although its phonology is the same as that of the source language and although most of its lexicon is Ngbandi, it is typically a pidgin in having a very limited vocabulary (from 500 to 1000 words in daily use) and virtually no inflection in its grammar;

tone plays a very limited grammatical role and only in the speech of those influenced by Ngbandi speakers, such as the Sangos and Yakomas.

Despite Sango's linguistic limitations, it is a symbol of Central African identity and is by far the preferred language of daily discourse in its capital of one-half million persons, Bangui. However, several Central Africans are active in legitimizing Sango with claims about its adequacy for indigenous culture and with efforts to increase its lexicon with Sango-based neologisms and words from regional languages. Nonetheless, French words occur in all varieties of Sango. The influence of French has increased since independence in vocabulary, grammar, and syntax, even among those with little education.

Although Sango was remarkably uniform as a *lingua franca*, it has become extremely variable as the vernacular of Bangui in all of its structures but exceptionally in its phonology. Contraction creates most of the word variants, as *twa* from *tongana* 'when, if,' resulting in many syllable and word forms that are strikingly different from those of indigenous languages: e.g., *tl* from *tî* 'of' with *l* carrying high tone with words beginning with *l*. Central African activists, however, are striving for a 'standard' (that is, normative or prescriptive) form of the language.

Practically all of the estimated 2 500 000 inhabitants of the Central African Republic speak Sango (according to the census of 1988, varying from 10 to 100%), and in 1994, it was the only language known by about 40–50% of Bangui's preschool non-Muslim children.

Bibliography

- Bouquiaux L (1976). *Contes de Tolé ou les avatars de l'aragne [sic] (République Centrafricaine): Contes recueillis*. Paris: CILF [Conseil International de la Langue Française].
- Bouquiaux L, Diki-Kidiri M & Kobozo J M (1978). *Dictionnaire sango-français et lexique français-sango*. Paris: SELAF.
- Calloc'h J (1911). *Vocabulaire français-sango et sango-français (langue commerciale de l'Oubangui-Chari) précédé d'un abrégé grammatical*. Paris: Paul Geuthner.

- Déchamps Wenezou M (1981). *Le français, le sango et les autres langues centrafricaines: Enquête sociolinguistique au quartier Boy-Rabe (Bangui, Centrafrique)*. Paris: SELAF.
- Diki-Kidiri M (1977). 'Développement du sango pour l'expression du monde moderne: Obstacles et possibilités.' In *Les relations entre les langues négro-africaines et la langue française, Dakar, 23–26 Mars 1976*. Paris: CILF [Conseil International de la Langue Française]. 717–728.
- Diki-Kidiri M (1977). *Le sango s'écrit aussi*. Paris: SELAF.
- Diki-Kidiri M (1986). 'Le sango dans la formation de la nation centrafricaine.' *Politique Africaine* 23, 83–99.
- Diki-Kidiri M (1995). 'Le sango.' In Boyd R (ed.) *Le système verbal dans les langues oubangiennes* [sic]. Munich/Newcastle: Lincom Europa. 141–164.
- Diki-Kidiri M (1998). *Dictionnaire orthographique du sängö*. Reading: BBA [Bûku tî Bêafrika].
- Diki-Kidiri M (1998). 'Le sango.' In Roulon-Doko P (ed.) *Les manières d' 'être' et les mots pour le dire dans les langues d'Afrique centrale*. Munich: Lincom Europa.
- Samarin W J (1967). *A grammar of Sango*. The Hague: Mouton.
- Samarin W J (1970). *Sango, langue de l'Afrique centrale*. Leiden: E. J. Brill.
- Samarin W J (1979). 'Simplification, pidginization, and language change.' In Hancock I F (ed.) *Readings in creole studies*. Ghent: E. Story-Scientia. 55–68.
- Samarin W J (1982). 'Colonization and pidginization on the Ubangi River.' *Journal of African Languages and Linguistics* 4, 1–42.
- Samarin W J (1982). 'Goals, roles, and language skills in colonizing central equatorial Africa.' *Anthropological Linguistics* 24, 410–422.
- Samarin W J (1984). 'The linguistic world of field colonialism.' *Language in Society* 13, 435–453.
- Samarin W J (1984/1985). 'Communication by Ubangian water and word.' *Sprache und Geschichte in Afrika* 6, 309–373.
- Samarin W J (1986). 'The source of Sango's "be."' *Journal of Pidgin and Creole Languages* 1(2), 205–223.
- Samarin W J (1989). 'Language in the colonization of central Africa, 1880–1900.' *Canadian Journal of African Studies* 23(2), 232–249.
- Samarin W J (1989). 'The colonial heritage of the Central African Republic: A linguistic perspective.' *The International Journal of African Historical Studies* 22(4), 697–711.
- Samarin W J (1994). 'The dynamics of morphotactic change in Sango.' In Moore K E, Peterson D A & Wentum C (eds.). *Proceedings of the Twentieth Anniversary Meeting of the Berkeley Linguistics Society, February 18–21, 1994* (Special session: Historical issues in African linguistics, February 18, 1994). Berkeley, CA: Berkeley Linguistics Society. 125–138.
- Samarin W J (1997). 'The creolization of pidgin morphophonology.' In Spears S K & Wynford D (eds.) *The structure and status of pidgins and creoles*. Amsterdam: John Benjamins. 175–210.
- Samarin W J (1998). 'The creation and critique of a Central African myth.' *Revue Française d' Histoire d'Outre-mer* 318(85), 55–81.
- Samarin W J (2000). 'The status of Sango in fact and fiction: on the one-hundredth anniversary of its conception.' In McWhorter H (ed.) *Current issues in pidgin and creole studies*. John Benjamins. 301–333.
- Samarin W J (2001). 'Explaining shift to Sango in Bangui.' In Nicolai R (ed.) *Leçons d'Afrique: Filiations, ruptures et reconstitution de langues. Un hommage à Gabriel Manessy*. Louvain/Paris: Peeters. 351–391.
- Samarin W J (2001). 'The past and present in marking futurity in Sango.' *Journal of Pidgin and Creole Languages* 16(1), 53–106.
- Samarin W J & Diki-Kidiri M (1983). 'Modernization in Sango.' In Fodor I & Hagège C (eds.) *Language reform: History and future* 3. Hamburg: Helmut Buske Verlag. 157–172.
- Samarin W J & Walker J A (1997). 'Sango phonology.' In Kaye A S (ed.) *Phonologies of Asia and Africa (including the Caucasus)* 2. Bloomington, IN: Eisenbrauns. 861–880.
- Saulnier P (1993). *Bangui chante: Anthologie du chant moderne en Afrique centrale*. Paris: L'Harmattan.
- Taber C R (1979). 'French loan-words in Sango: The motivation of lexical borrowing.' In Hancock I F (ed.) *Readings in creole studies*. Ghent: E. Story-Scientia. 189–197.

Sanskrit

J L Brockington, University of Edinburgh,
Edinburgh, UK

© 2006 Elsevier Ltd. All rights reserved.

The Sanskrit language – one of the oldest of the Indo-European group to possess a substantial literature – has particular interest for linguists because of the circumstances of its becoming known to Western

scholars and the stimulus thus given to historical linguistics. It has also been of enormous and continuing importance as the classical language of Indian culture and the sacred language of Hinduism.

Origin and History

Sanskrit, in its older form of Vedic Sanskrit (or simply Vedic), was brought into the northwest of India by

the Āryans some time in the second half of the second millennium BC and was at that period relatively little differentiated from its nearest relation within the Indo-European group, Avestan in the Iranian family of languages (these two being the oldest recorded within the Indo-Iranian branch of Indo-European). From there, it spread to the rest of North India as the Āryans enlarged the area that they occupied, developing into the classical form of the language, which subsequently became fixed as the learned language of culture and religion throughout the subcontinent, while the spoken language developed into the various Prākṛits. There is ample evidence of rapid evolution during the Vedic period, with the language of the latest phase, attested for example in the Upanisads, showing considerable grammatical simplification from that of the earliest hymns. The later Vedic is, in broad terms, the form of the language that Pāṇini described with such exactness in his grammar around the fourth century BC, thereby creating – no doubt unintentionally – an absolute standard for the language thereafter; his work is clearly the culmination of a long grammatical tradition, based on concern to preserve the Vedas unaltered (hence the stress on phonetics), and is itself intended for memorization and oral transmission, as its brevity indicates.

This standardization was not as universal as has sometimes been represented (nor was the preceding Vedic a unified language, for it exhibits features only explicable as coming from slightly differing dialects, and classical Sanskrit is based on a more eastern dialect than the one attested in the *R̥gveda*), and it has come to be recognized that, for example, the two Sanskrit epics exhibit systematic divergences from the language described by Pāṇini and represent a distinct epic dialect. However, with the growth of classical Sanskrit literature (mainly within the period from the fourth to the tenth centuries AD, when Sanskrit was clearly no longer a natural language), Pāṇini's description was regarded as prescriptive and followed to the letter, although the spirit was less closely observed (as shown by the tendency to longer and longer compounds and to nominal constructions and the like).

The earliest record of the language is contained in the hymns of the *R̥gveda*, which belong to around 1200–1000 BC, but they were not committed to writing until a much later period because of their sacred character, for the Indian tradition has always placed greater emphasis on oral tradition than on written texts. In fact, the earliest dated record in Sanskrit is an inscription of 150 AD, significantly later than the use of Prākṛit by the Buddhist ruler Aśoka for his inscriptions in the third century BC. Early inscriptions used one of two scripts: the Kharoṣṭhī, deriving from

the Aramaic script used in Achaemenid Iran, and the Brāhmī, less certainly deriving from a North Semitic script. The latter evolved into the Nāgarī family of scripts, to which the Devanāgarī script now usually used for Sanskrit belongs, although before the twentieth-century manuscripts were normally written in the local script.

Characteristics

Any analysis of Sanskrit syntax must take account of the shift from the natural language of the Vedic and epic forms of Sanskrit to the learned language of the classical literature, which selectively exploits certain features of Pāṇini's description. Whereas the older forms of the language show frequent use of nominal compounds of two or three members and Pāṇini's grammar describes their formation in great detail (but in terms of their analysis into types: *dvandva*, *bahuvrīhi*, *tatpuruṣa*), classical literature is marked by a predilection for longer compounds, consisting in some styles of writing of 20 or more members. Another common feature, inherited from the Indo-European background but found much more extensively in the classical language, is the use of nominal sentences involving the juxtaposition of the subject and a nonverbal predicate. The frequent use of the past participle passive as a verbal equivalent leads to a preference for passive constructions, in a way typical of the Prākṛits. Use of the absolutive becomes in the classical language a common means to form complex sentences by indicating actions occurring before that of the main verb; again the effect is a reduction in finite verbal forms. The usual sentence order is subject, object, verb; however, this is so commonly modified for emphasis (with initial and final positions in the sentence or verse-line carrying most emphasis) that Sanskrit word order is often regarded as being free. In vocabulary, the freeing from the affective connotations of a natural language brought a striking enlargement of the range of synonyms, skillfully exploited in much of the classical literature to produce rich sound effects.

In its morphology, Sanskrit is broadly comparable to Greek or Latin, though somewhat more complex. In both the nominal and verbal systems the dual is obligatory for all twos, not just pairs. The nominal system employs eight cases (seven according to the Indian reckoning, which regards the vocative as a form of the stem), three numbers, and three genders (masculine, feminine, neuter). Unlike other Indo-European languages, Sanskrit lacks a developed series of prepositions, and the relatively few adverbial formations used to define case relationships more exactly tend to be placed after the noun. The use of

vrddhi (IE strengthened grade) to form derivatives from nominal stems is a notable feature. The verb has two voices, active and middle, their functions well distinguished by the Sanskrit terms for them: *parasmaipada* ‘word for another’ and *ātmanepada* ‘word for oneself’; it also has five moods (injunctive, imperative, subjunctive, optative, and precative) in the Vedic, somewhat simplified in the classical language. Prepositional affixes to the verb may in Vedic be separated from the verb but in the classical language must be prefixed to it (there is a comparable development between Homeric and classical Greek). There is both an ordinary sigmatic future and a periphrastic future (formed through a specialized use of the agent noun), several aorist formations (principally a sigmatic aorist and a root aorist), and a perfect normally formed with a reduplicated stem; these are comparable to the equivalent tenses in Greek or Latin. The augment is prefixed to several past tenses: imperfect, aorist, pluperfect, and conditional. Verbal roots are divided by the Sanskrit grammarians into 10 classes: six athematic and four thematic. A distinctive feature of the verbal system is the employment of secondary conjugations with specific meanings: causative, intensive, and desiderative. Historically, the passive is also such a secondary conjugation, formed by adding the middle endings to a modified root. The Vedic language is marked by rather greater grammatical complexity with, most notably, a whole range of case forms from nouns functioning as infinitives, which are reduced to a single infinitive in the classical language. It also possessed a pitch accent that had died out by the time of the classical language.

Phonetically Sanskrit is marked by a number of innovations by comparison with other Indo-European languages of comparable age. It is also notable for the concern with phonetics of its own grammarians (exemplified by the fact that the alphabet is arranged according to the organ of articulation, with vowels preceding consonants) and the precision of their descriptions. On the one hand, Sanskrit has collapsed the three Indo-European vowels *a*, *e*, and *o* into *a*, and on the other it has introduced a complete new class of consonants, that of the retroflex consonants, mainly under the influence of one of the other language groups already present in India, either Dravidian or Munda, although in some instances the retroflex consonants probably arose through internal phonetic developments in relation to *ṣ* and *r*. The most widely known feature is that of *saṃdhi* ‘junction,’ the process of phonetic assimilation of contiguous sounds at the junctures between both words and their component parts (external and internal *saṃdhi*).

Sample Sentence

| | | | | |
|----------------|-----------------|-----------------|-----------------|--------------|
| <i>teṣāṃ</i> | <i>khālū</i> | <i>eṣāṃ</i> | <i>bhūṭānāṃ</i> | <i>trīṇy</i> |
| /teṣa:ṇ | khəlv | eṣa:ṇ | bhu:ta:na:ṇ | tri:ṇy |
| <i>eva</i> | <i>bījāni</i> | <i>bhavanti</i> | <i>aṇḍajam</i> | |
| evə | bi:ja:ni | bhəvənti | əṇḍəjəṇ | |
| <i>jīvajam</i> | <i>udbhijam</i> | <i>iti</i> | | |
| ji:vəjəṃ | udbhijəṇ | iti/ | | |

‘Living beings here have just three origins [literally ‘Assuredly of these living beings are/come into being indeed three seeds’]: being born from an egg or live-born or produced from a sprout.’

This simple sentence (from *Chāndogya Upaniṣad* 6.3.1) exemplifies several of the features that are taken to extremes in the classical language. There is the avoidance of a transitive construction (although here the verb, *bhavanti*, is expressed, whereas later such a copula is normally suppressed), the employment of compounds, and the liking for etymological figures (the latter two combined in the three compounds ending in the adjectival form *-ja*, coming from *√jan* ‘to be born,’ while the use of cognates is exemplified by *bhavanti* third pl present indicative and *bhūta* past participle passive from *√bhū* ‘to become’). The use of *iti* may also be noted – here to function as the equivalent of the colon in the translation, more usually to perform the function of quotation marks, to mark off a passage in direct speech from the sentence in which it is embedded (an idiom probably calqued on the Dravidian); Sanskrit has no method of indicating indirect speech.

Role and Influence in Indian Culture

As is implicit in some of the statements above, it is clear that throughout the main period of its use as a literary language Sanskrit was not the first language of its users, who in North India would have been native speakers of one of the Prākṛits deriving from Sanskrit (used here in its widest sense of the group of OIA dialects) or even of the next stage of MIA, the Apabhraṃśas, and in South India were speakers of one of the Dravidian languages (which have been influenced to varying degrees in their vocabulary by Sanskrit). The prestige attaching to its use for the Vedas, the authoritative scriptures for Hindus, resulted in its being regarded as the only language fit for use in the major rituals of brahmanical Hinduism, a role that to a limited extent it retains to this day. This was undoubtedly the reason why the Purāṇas and the many popular texts related to them were composed (from the fourth century to as late as the nineteenth century) in a form of Sanskrit that is greatly indebted to the epics for its linguistic and metrical expression, while similarly Mahāyāna Buddhism employed the so-called Buddhist

Hybrid Sanskrit (essentially a Sanskritization of MIA). Sanskrit has therefore been a dominant influence on the development of the languages in both the MIA and NIA phases, supplying much of the religious vocabulary in the form of direct loans, over and above the large proportion of the vocabulary descended from Sanskrit.

Sanskrit and the West

First acquaintance with Sanskrit by Western scholars came even before the period of British rule. Sir William Jones's famous discourse in 1786 to the Asiatic Society in Calcutta on the affinity of Sanskrit with Greek, Latin, and the other languages now known as Indo-European was not the first notice of such connection, which had been proposed two centuries earlier by Thomas Stevens (in 1583) and Fillipo Sassetti (in 1585). However, Jones's eminence ensured it a much wider audience than before, and this was in a signifi-

cant sense the start of the discipline of comparative philology, whereas the appreciation before long of the achievements of the early Indian grammarians was an important stimulant to the development of modern linguistics, which has paid them the compliment of borrowing a number of their terms, such as *saṃdhi*.

Bibliography

- Burrow T (1973). *The Sanskrit language* (3rd edn.). London: Faber and Faber.
- Cardona G (1988). *Pāṇini: his work and its traditions*. Delhi: Motilal Banarsidass.
- Coulson M (1992). *Sanskrit: an introduction to the classical language* (2nd edn.). London: Hodder and Stoughton.
- Gonda J (1971). *Old Indian*. Handbuch der Orientalistik, 2. Abt., 1. Bd., 1. Abschnitt. Leiden-Cologne: E. Brill.
- Scharfe H (1977). *Grammatical literature*. History of Indian Literature, vol. V. fasc. 2. Wiesbaden: Harrassowitz.

Santali

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

Santali (𑒧𑒻𑒟𑒱𑒪𑒲), a member of the North Munda (Kherwarian) subgroup of the Munda family within the Austroasiatic linguistic phylum, is spoken by between 5 million and 7 million people across several states in eastern and central India. The most compact area of Santal settlement is in the Sadar subdivision of Bunkura, the Jhargram subdivision of Midnapur, and Purulia in West Bengal; south of Bhagalpur and Monghyr, in the Santal Parganas, Hazaribagh and Dhalbhum in Bihar, and the newly formed tribal-dominant state of Jharkhand; and Baleshwar, Mayurbhanj, and Keonjhar in Orissa. In Bangladesh, the Santals are found mainly in Rajsahi, Rangpur, and the Chittagong Hill tracts (Ghosh, 1994: 3).

Santali is characterized by a split into at least a northern and southern dialect sphere, with slightly different sets of phonemes (Southern Santali has six phonemic vowels, in contrast with eight or nine in Northern Santali), different lexical items, and to a certain degree, variable morphology as well (e.g., the *-ič*:-*rən* singular:plural opposition in animate genitive case markers).

There is a degree of laryngeal tension (phonation type) associated with certain Santali vowels. This

gives Santali and the closely related Mundari their characteristic sound and differentiates these two languages from other languages of the region. Instrumental studies are needed to determine the exact phonetic characteristics of this. In addition, a wide range of vowel combinations may be found in Santali; this tendency finds an extreme expression in words such as *kəəəəə* 'he will ask for him'.

The following statements can be made regarding the consonantism of Santali: retroflexion, while attested, is less developed in Santali (and Munda) generally than in Indo-Aryan or Dravidian languages. Further, in coda position, there is a characteristic use of so-called checked consonants, ranging in articulation from preglottalized to unreleased. Examples of checked consonants in final position in Santali include *seč* 'towards', *rit* 'grind', *selep* 'antelope', and *dak* 'water' (Bodding, 1923: 79); before vowels (generally speaking), these consonants alternate with voiced stops, as in *dal-aka-t-ko-a-e* 'he has beat them' vs. *dal-aka-d-e-a-e* 'he has beaten him'. Santali also makes use of prenasalized stops in a number of words as well: *k^bokṇdo* 'ill conditioned', *məñ^blə* 'fourth of six brothers', *məñ^bj* 'beautiful', *o^tṇgao* 'to steady on', *ganḍke* 'log', *onḍga* 'ogre', *b^bosṇdo* 'slovenly', *bermbak* 'incorrectly', and *teṇga* 'stick'; also *k^bəṇduṇ* ~ *k^bəṇduṇ* 'deep', *kəñ^jɛ* (~*kəñ^jɛ*) 'crooked', and *d^baṇga* (~*dhəṇgra/i*) 'strapping' (Bodding, 1923: 36ff.).

Santali has a complicated demonstrative system (Zide, 1972). Its basic three-way system is a straight-forward proximal, distal, remote system in animate (-i/kin/ko) and inanimate forms (-a/-akin/-ako), as shown in the following examples (ANIM, animate; INAN, inanimate; SG, singular; DL, dual; PL, plural):

(1a) Proximal:

| | SG | DL | PL |
|-------|--------|-----------|-----------|
| ANIM: | nui | nukin | noko/nuku |
| INAN: | noa | noakin | noako |
| | 'this' | 'these 2' | 'these' |

(1b) Distal:

| | SG | DL | PL |
|-------|--------|-----------|-----------|
| ANIM: | uni | unkin | onko/unku |
| INAN: | ona | onakin | onako |
| | 'that' | 'those 2' | 'those' |

(1c) Remote:

| | SG | DL | PL |
|-------|---------------|--------------------|--------|
| ANIM: | hani | hankin | hanko |
| INAN: | hana | hanakin | hanako |
| | 'that yonder' | 'those (2) yonder' | |

Alongside these are intensive forms (Example (2); marked by infixation of -k'-), 'just' forms (Examples (3a) and (3b); marked by a shift of (o/u>)-i-), as well as forms adding connotations of 'things seen' and 'things heard' (Examples (4a)–(4c)):

(2) Intensives:

| | |
|--------|----------------------|
| nuk'ui | 'this very one' |
| nik'i | 'just this very one' |
| nok'oy | 'this very thing' |

(3a) 'Just' proximal:

| | SG | DL | PL |
|-------|-------------|----------------|--------------|
| ANIM: | nii | nikin | neko/niku |
| INAN: | nia | niakin | niako |
| | 'just this' | 'just these 2' | 'just these' |

(3b) 'Just' distal:

| | SG | DL | PL |
|-------|-------------|------------------|-----------|
| ANIM: | ini | inkin | enko/inku |
| INAN: | ina | inakin | inako |
| | 'just that' | 'just those (2)' | |

(4a) 'Seen' distal:

| | SG | DL | PL |
|--|-------------|------------------|-------|
| | ɔne | ɔnekin | ɔneko |
| | 'that seen' | 'those (2) seen' | |

(4b) 'Seen' remote:

| | SG | DL | PL |
|--|--------------------|-------------------------|--------|
| | hane | hanekin | haneko |
| | 'that yonder seen' | 'those (2) yonder seen' | |

(4c) 'Heard' distal:

| | SG | DL | PL |
|--|--------------|-------------------|-------|
| | ɔte | ɔtekin | ɔteko |
| | 'that heard' | 'those (2) heard' | |

Verbs as a lexical category in Santali, and indeed in Munda languages generally speaking, are not easily or rigorously defined in opposition to nouns (Bhat, 1997; Bhattacharya, 1975; Cust, 1878; Pinnow, 1966). As seen in the following examples (Ghosh, 1994: 21), one and the same root may be used as a noun, as a modifier (adjective/participle), and as a predicate/verb. Even a noun root such as 'house' may be used verbally with verbal inflection (ASP, aspect; TR, transitive; FIN, finite):

| | | |
|------|---------|-----------------|
| (5a) | kombro | kombro mɛɾɔm |
| | thief | stolen goat |
| | 'thief' | 'a stolen goat' |

| | | |
|------|------------------------|--------------------|
| (5b) | mɛɾɔm-ko | kombro-ke-d-e-a |
| | goat-PL | steal-ASP-TR-3-FIN |
| | 'They stole the goat'. | |

| | |
|------|--------------------|
| (5c) | oɾak-ke-d-a-e |
| | house-ASP-TR-FIN-3 |
| | 'He made a house'. |

The default position for subject agreement clitics is in immediately preverbal position in Santali. Note in the following examples (Bodding, 1929a: 58, 60, 208) that this is true even if the element appearing in this position is an overt subject (or object) pronoun (1, first person; INTR, intransitive; 2, second person; LOC, locative; ALL, allative):

| | | |
|------|---------------------------------|----------------------|
| (6a) | Kumbrɔbad-te-ko | əgu-ke-'t-le-a |
| | K-LOC/ALL-PL | bring-ASP-TR-1PL-FIN |
| | 'They brought us to Kumbrabad'. | |

| | | | |
|------|------------------|-------|-------------|
| (6b) | hɛ | iñ-iñ | cala'k-a |
| | yes | I-1 | go.INTR-FIN |
| | 'Yes I will go'. | | |

| | | | |
|------|-------------------|-------|-----------|
| (6c) | iñ | am-iñ | ñɛl-mɛ-a |
| | I | you-1 | see-2-FIN |
| | 'I will see you'. | | |

A wide range of arguments or referents may be encoded within the Santali verbal complex. This includes subjects, direct objects, indirect objects, benefactives, and possessors of subjects or objects. Note that Santali is doubly unusual in its system of possessor indexing: it takes a special series of possessive inflection, and this pattern of referent indexing does not reflect a process of 'raising' (to argument/term status of this logical modifier/operator), as a verb in Santali may encode both its logical argument and a possessor of that argument simultaneously, as in Example 7d). Examples (7a)–(7d) are from Bodding (1929a: 212, 1923: 22, 21–22, 209), respectively, and Example (7e) is from Ghosh (1994: 65) (NEG, negation; ANT, anterior; BEN, benefactive; POSS, possessor):

| | | |
|------|---------------------------|--------------------|
| (7a) | ba-ko | sap'-le-d-e-a |
| | NEG-PL | catch-ANT-TR-3-FIN |
| | 'They did not catch him'. | |

- (7b) im-əñ-me
give-1-2
'give me'
- (7c) dul-a-ñ-me
pour.out-BEN-1-2
'Pour out for me'.
- (7d) sukri-ko gɔ'c-ke-d-e-tiñ-a
pig-PL die-ASP-TR-3-1.POSS-FIN
'They killed my pig'.
- (7e) hɔpɔn-e hɛč'-en-tiñ-a
son-3 come-PAST.INTR-1.POSS-FIN
'My son came'.

The Santali language has been written in at least five alphabets, depending on the locale of production and the purpose of the written material. There have been Santali publications in Devanagari (Hindi), Oriya, Bengali, and Roman and in the *ol čemət* ~ *ol čiki* script of indigenous origin (Zide, 2000: 8). An ever-growing body of literature has appeared in Santali, and the language is used on a limited basis in other media (e.g., shortwave radio broadcasts).

Bibliography

- Anderson G D S (2001). 'Santali.' In Garry J & Rubino C (eds.) *Facts about the world's languages*. Bronx, NY: H. W. Wilson.
- Anonymous (1929). *Tea districts Labour Association language handbook: Santhali*. Calcutta: Catholic Orphan Press.
- Bhat D N S (1997). 'Noun-verb distinction in Munda languages.' In Abbi A (ed.) *Languages of tribal and indigenous peoples of India: the ethnic space*. Delhi: Motilal Banarsidass. 227–251.
- Bodding P O (1923). *Materials for a Santali grammar (mostly phonetic)*. Benegaria: Santal Mission Press.
- Bodding P O (1929a). *Materials for a Santali grammar (mostly morphological)*. Benegaria: Santal Mission Press.
- Bodding P O (1929b). *A Santali grammar for beginners*. Benegaria: Santal Mission Press.
- Bodding P O (1925–1929). *Santal folk tales* (3 vols). Oslo: Instituttet for Sammenligende Kulturforsning.
- Bodding P O (1929–1936). *A Santal dictionary* (5 vols). Oslo: Norwegian Academy of Science and Letters.
- Campbell A (1899–1902). *A Santali-English dictionary*. (3 parts). Pokhuria: Santal Mission Press.
- Cole F T (1906). *Santali primer*. Pokhuria: Santal Mission Press.
- Cust R N (1878). *A sketch of the modern languages of the East Indies*. London: Trübner.
- Ghosh A (1994). *Santali: a look into Santali morphology*. New Delhi: Gyan Publishing House.
- Heuman E (1892). 'Grammatisk studie öfver santal-språket.' *Det kgl. Danske Videnskabernes Selskabs forhandling*, Copenhagen, 148–220.
- Kisku P C & Soren K R (1951). *Santali śabda kosh*. Deoghar: Santal Paharia Seva Mandal. [In Santali.]
- Lekomtsev J K (1968). 'Nekotorye kharakternye cherty santal'skogo predlozhenija.' In *Jazyki Indii, Pakistana, Nepala i Sejlona*. Moscow: Nauka. 311–320.
- Lekomtsev J K (1975). 'Nekotorye nabludenija nad morfologiej Santali.' In Elizarenkova T I (ed.) *Očerki po fonologii vostochnyx jazykov*. Moscow: Nauka. 178–205.
- Macphail R M (1954). *Campbell's English-Santali dictionary* (3rd edn.). Benegaria: Santal Mission Press.
- Macphail R M (1964). *An introduction to Santali*. Benegaria: Santal Mission Press.
- Murmu R (1976). *Ranar: a Santali grammar in Santali (Ol script work)*. Baripada: Murmu.
- Neukom L (2001). *Santali. Languages of the world/materials*, 323. München: Lincom.
- Pal A (1980a). 'How action is expressed in Hor Parsi or Santali language.' *Journal of the Asiatic Society of Calcutta* 22/3–4, 151–162.
- Pal A (1980b). 'The structure of qualifying words in Santali.' *Journal of the Asiatic Society of Calcutta* 22/3–4, 191–196.
- Phillips J (1852). *An introduction to the Santali language*. Calcutta: Calcutta School Book Society's Press.
- Sahu D (1962; 1968). *Santali pravesika* (2 vols). [In Santali.] Dengara, Bihar: Prapti-sthana Abhirama Prakasa.
- Skrefsrud L O (1873). *A grammar of the Santhali language*. Benares: Medical Hall Press.
- Zide N H (1958). 'Final stops in Korku and Santali.' *Indian Linguistics* 19, 44–48.
- Zide N H (1967). 'The Santali Ol Cemet script.' In *Languages and areas: studies presented to George V. Bobrinskoy*. Chicago: University of Chicago. 180–189.
- Zide N H (1972). 'A Munda demonstrative system: Santali.' In Barrau J et al. (eds.) *Langues et techniques, nature et société I (papers for A. Haudricourt)*. Paris: Klincksieck. 267–274.
- Zide N H (2000). 'Three Munda scripts.' *Linguistics in the Tibeto-Burman Area* 22, 199–232.

Scots

C Macafee, University of Aberdeen, Aberdeen, UK

© 2006 Elsevier Ltd. All rights reserved.

Scots comprises a group of dialects spoken in Lowland Scotland, forming a continuum with Northern English dialects, and in Orkney and Shetland. Scots was taken to Ireland in the 17th century and survived intact there in parts of Northern Ireland and County Donegal. Individual linguistic items also survive in Canada, the United States, and New Zealand. The UK government has given Scots *de facto* recognition by listing it in the European Charter for Regional or Minority Languages, albeit in a section that implies no specific commitments. Ulster Scots is recognized as a language and is given financial support by the UK and Irish governments, under the terms of an intergovernmental agreement.

Scots is in a sociolinguistic continuum with Scottish Standard English (SSE), and is often intermingled with it in practice. Speakers who maintain a rich and focused variety of Scots are found mainly in rural areas and small towns outside the central industrial belt. In the North-East and the Northern Isles, it is not uncommon for professional people to code switch sharply between Scots and English. There are concerns, however, that this ability may not be continuing among younger people. In the most densely populated areas of Scotland, Scots is spoken mainly lower down the social scale.

A question on Scots (parallel to that on Gaelic) was considered for inclusion in the 2001 Census, but it was concluded that a valid and reliable question could not be framed. The question-testing exercises suggested that almost a third of the Scottish population might call themselves speakers of (a dialect of) Scots, but, in the absence of a popularly accepted terminology, the responses were affected by the form of the question. Urban dialects, which are thinnest in traditional vocabulary, are least likely to be identified with the historical language or to be dignified with the name 'Scots,' and are most likely to be perceived as slang. There is some teaching and research on Scots at university level in Scotland and Northern Ireland, and at the Royal Scottish Academy of Music and Drama in Glasgow. In Northern Ireland, the Ulster Scots Agency supports cultural and linguistic projects, but attempts to promote a hastily modernized version have been met unsympathetically by speakers and nonspeakers alike. In Scotland, a modernized form of Scots ('Lallans') is promoted by the Scots Language Society.

Scots, like Standard English, is descended from the Anglian dialect of Old English. It is not clear to what extent this variety, established south of the Forth in the 7th century, was the ancestor of Scots, and to what extent it was swamped by a later influx of speakers, mainly from Yorkshire. These were followers of Anglo-Norman mercenaries and monastics, invited in by the Scottish crown in large numbers in the 12th century. Anglian spread throughout the Lowlands with the establishment of urban settlements (burghs) and of Norman administration in church and civil life. As a result, Lowland Scots is rich in loans from Old Norse, mostly shared with Northern English dialects. In Caithness, Orkney, and Shetland, where there was extensive Norse settlement, Scots replaced an extinct Scandinavian language, Norn (spoken in Shetland as late as the 19th century). Some loans survive, e.g., *bonxie* 'the great skua' and *moorit* 'brown (of sheep)'. Apart from place-names, Scots texts begin in the 14th century with glosses and phrases in Latin and French documents, and a few legal papers in a Scots are still very similar to Northern English. The long narrative poem, 'The Bruce,' by John Barbour, was written around 1375, but surviving manuscripts are from the early 15th century.

By the 15th century, the orthography of Scots was distinctive. The spellings <ee, oo, ea> were avoided. For /x/, <ch> was used, as in *nicht* 'night'. Words such as 'little' ended in <ill>, thus *litill*. As in Northern English, <quh> was used rather than <wh>, and <sch> rather than <sh>, and there was interchangeability among <u, v, w>, but Scots was unusual in using <w> initially, as in *wþ* 'up'. Distinctively, Scots used <l3> for French- and Gaelic-derived /l/, as in *tulzie* 'a struggle', and <n3> for similarly derived /n/, as in *cunze* 'a coin'. As the retention of these consonants in Early Scots illustrates, the influence of French was independent of its influence on English, and the same was true of Latin, so that loans from these languages further differentiated Scots, as did loans from Middle Dutch (through immigration to the burghs, and fishing and trade contacts), and, of course, loans from Gaelic. The long-term Gaelic influence on vocabulary is, however, unexpectedly small. It would appear that the transition from Gaelic to Scots was effected with some completeness. Gaelic has had phonological influences, e.g., North-Eastern /f/ for /hw/ (earlier /xw/), as in *fa* 'who', and grammatical influences, e.g., emphatic use of reflexives, as in 'It's yourself!'

In the mid-15th century, Early Scots gave way to Middle Scots (both being stages of Older Scots).

Middle Scots saw the spread (from Northern England) of ‘i-digraph’ spellings for stressed monophthongs—thus <ai>, as in <haim> for <hame> ‘home’; <ei>, as in <leid> for <lede> ‘lead’ (noun or verb); <oi>, as in <rois> for <rose> (the flower); and <ui>, as in <muin> for <mune> ‘moon’. As generally north of the Humber, words such as *hame* ‘home’ have retained a front development of Old English *ā*, raised by the ‘Great Vowel Shift,’ and words such as *mune* ‘moon’ have a fronted development of Old English *ō*.

Scots and its dialects are characterized by numerous conditioned sound changes. To mention just one dating from this time, there was a vocalization of /l/ following the short vowels /a, o/ and, less regularly, /u/. Examples include *ba* ‘ball’, pronounced /bɔ/ (Modern Central dialect) or /ba/ (Northern dialect); *gowd* /gɔud/ ‘gold’; and *fu* /fu/ or *full* /fɔl/ ‘full’. A standard based on Edinburgh was emerging when Scots was replaced in formal use by Standard English in the late 16th century, largely as a result of the printing press and of the Reformation of 1560, which introduced the Bible in English translation. Standard English became the speech of the ruling class following the Union of the Crowns in 1603, and of the professions following the Union of the Parliaments in 1707. Insistence that schoolteachers use English began with national inspection in 1845.

Literature and historical documents of all kinds exist from the Older Scots period, including the work of ‘makars’ such as William Dunbar, Robert Henryson, and Gavin Douglas. Well-known writers from the modern period include Robert Burns, John Galt, Walter Scott, and Robert Louis Stevenson. Folk tales and songs (including the ‘muckle sangs,’ or Child ballads) have been extensively collected. The 20th-century folk revival has been very important for the continuation of elevated styles of Scots.

Language revival has been a spur to literature, including the poetry of Hugh MacDiarmid (Christopher Grieve), the songs of Hamish Henderson, and drama (original and in translation). Notable work continues also in local vernaculars – for instance, in the poetry of Sheena Blackhall in North-East Scots (‘the Doric’). The urban dialects, being thinner, are more widely accessible, as witness the work of Liz Lochhead, Tom Leonard (both Glasgow), and Irvine Welsh (Edinburgh). Scots, when it is occasionally used in television and radio scripts, is generally thin, for the same reason.

Modern Scots (from 1700 on) has preserved little from Older Scots in orthography, apart from <ch> and <ui> spellings. The 18th-century introduction of <oo> for /ø/, as in /tøm, tɪm/ ‘empty’, though now replaced by <uCe, ui>, still causes some confusion

in spelling and pronunciation in historical contexts. Apostrophes for ‘missing’ letters, such as for /d/ after /n/, as in *en* ‘end’, or /d/ after /l/, as in *aul* ‘old’, and for final <th>, as in *wi* ‘with’, were also a feature of Modern Scots, but have largely been dropped as the result of spelling reform. The following example is of Modern Central Scots:

The laddies gets hame fae the schule afore the wee lass an they wint their tea the meenit they’re in the hoose. Bit it’ll no dae thaim a scart o herm tae thole their hunger a while langer.

ðɪ 'ladez ɡɪts 'hem fe ðɪ 'skɪl ʌ'for ðɪ 'wi 'las n, ðe 'wɪnt ðɪr 'tɪ ðɪ 'mɪnɪt ðɪr ɪn ðɪ 'hus. bɪt ɪtɪl, 'no 'de ðʌm ʌ 'skart ɪ 'herm te 'θol ðɪr 'hʌŋɪr ʌ 'hwəɪl 'lʌŋɪr.

Scots has minor differences in grammar from English dialects; for instance, there is a double system of concord in the present tense of verbs: if the subject is a personal pronoun adjacent to the verb, the verb is inflected only in the third-person singular (and second-person singular, where it is preserved), otherwise in all persons and numbers. The dialects of Orkney and Shetland preserve the second-person singular, as in *du* ‘thou’. A new second-person plural form, *youse*, has spread from Ulster to Glasgow. In the demonstrative system, *yon* (or *thon*) expresses a greater distance than does ‘that’ or *thae* ‘those’.

The vowel length of stressed monophthongs (other than /ɪ, ʌ/, which are always short) is determined by the Scottish Vowel-Length Rule, with long vowels before the voiced fricatives /v, ð, z, ʒ/ and /r/, and morpheme finally, thus *agreed* long but *greed* short (likewise in SSE for a narrower range of vowels). Old English *ū* remains /u/, as in *about* ‘about’ and *hoose* ‘house’. Old English *ō* fronted to /ø/ remains in conservative dialects (mainly Orkney and Shetland), as in ‘do’ (from Old French), ‘use’ and *schule* ‘school’. In North-East Scots, this unrounded already in Middle Scots to /i/, thus *dee*, *meen*, and *eese*. In Modern Central Scots, the vowel unrounds in the long environments of the Scottish Vowel-Length Rule to /e/, thus *yaise* (verb), and to /ɪ/ elsewhere, thus *yis* (noun). In some dialects, this /ɪ/ from earlier /ø/ remains separate from the reflex of Old English /i/, as in ‘bring’, which is lower and more central, as [ɛ̃]. This may be the source of the similar realization of /ɪ/ in New Zealand English.

Old English *ī* has split in Modern Scots into /aɪ/ in the long environments of the Scottish Vowel-Length Rule, as in ‘five’ and ‘why’, and into /əɪ/ otherwise, as in ‘while’. (The similar allophony in Canadian English may owe something to Scots or to SSE.) This /əɪ/ merges with the reflex of *ai* word finally, as in *cley* ‘clay’, and in Anglo-Norman *ui*, as in

bile 'boil'. The unrounding of *ü* to /*ʌ*/ is complete: there is no /*ʊ*/. Thus 'push' and 'pull' have /*ʌ*/. As a consequence of the Scottish Vowel-Length Rule, the short vowel /*o*/, as in 'lot', *tod* 'fox' has merged in Central Scots with the long vowel /*o*/, as in *thole* 'endure' (from Old English *ō* by Open Syllable Lengthening).

The vowel phonology of SSE is largely based on the vowel system of Central Scots, but with the lexical incidence of Standard English. In the circumstances of the 18th century, with limited access to native speakers of English, a number of interdialectal features became fixed in SSE. Since there was no /*ʊ*/, these words were assigned to /*u*/, thus 'pool' and 'pull' are homophones in SSE. (In Central Scots, they are /*pɪl*/ and /*pu*/ or /*pʌl*/.) Similarly, since there was no /*ɒ*/, words such as 'cot' were assigned to Central Scots /*ɔ*/, as in *lauch* 'laugh' and *saut* 'salt'. Thus 'cot' and 'caught' are homophones in SSE. (In Central Scots, they are /*kot*/ and /*kɔxt*/; in Northern Scots, they are /*kɔt*/ and /*kaxt*/.) The similar absence of distinction between the vowels of 'cot' and 'caught' in some North American varieties may owe something to SSE. The lower and more central realization of /*ɪ*/ in Scots has left a residue of words in SSE with a phoneme /*ɛ*/, thus /*nɛvər*/ 'never', contrasting with /*rɪvər*/ 'river' and /*sɛvər*/ 'sever'.

Scots has a large shared vocabulary with English (for example, 'the,' 'cat,' and 'tell') and also much distinctive vocabulary, as in the Old English survivals *neep* 'turnip' and *een* 'eyes', the coinages *gully* 'knife' and *tapsalteerie* 'head-over-heels', Old Norse *nout* 'cattle' and *skellie* 'squint', Middle Dutch *craig* 'the neck' and *redd* 'clear up', Anglo-Norman *leal* 'loyal' and *hurcheon* 'hedghehog', Latin

stravaig 'roam', Central French *Hogmanay* 'New Year's Eve', and Gaelic *corrie-(fistit)* 'left-(handed)' and *sonse* 'prosperity'.

Bibliography

- Corbett J *et al.* (eds.) (2003). *The Edinburgh companion to Scots*. Edinburgh: Edinburgh University Press.
 Görlach M (2002). *A textual history of Scots*. Heidelberg: C. Winter.
 Jones C (ed.) (1997). *The Edinburgh history of the Scots language*. Edinburgh: Edinburgh University Press.
 Macafee C & Aitken A J (2002). 'A history of Scots to 1700.' In Craigie W *et al.* (eds.) *A dictionary of the older Scottish tongue*, vol. XII. Oxford: Oxford University Press.

Relevant Websites

- <http://www.unibg.it/anglistica/slin/scot-bib.htm> – Website of Università degli Studi di Bergamo, where 'A selected classified bibliography of the Scots language,' compiled by Caroline Macafee, is maintained.
<http://www.eng.helsinki.fi> – Website of the Department of English at the Faculty of Arts, University of Helsinki, where the Helsinki Corpus of Older Scots can be accessed.
<http://www.sldl.org.uk> – Website providing access to numerous Scottish Language Dictionaries (formerly the Scottish National Dictionary Association website).
<http://www2.arts.gla.ac.uk> – Website of the Association for Scottish Literary Studies, providing access to the Corpus of Modern Scots Texts.
<http://www.pkc.gov.uk> – Website for Perthshire and for Perth/Kinross Council, where the Scots Language Resource Centre can be accessed.

Scots Gaelic

K MacKinnon, University of Herefordshire,
Black Isle, UK

© 2006 Elsevier Ltd. All rights reserved.

Origins and Early History

The Scottish people originated with Gaelic-speaking incomers from northeastern Ulster who settled in the northwestern coastlands and islands of Caledonia in the later 5th century and subsequently relocated their kingdom of Dal Riata from Ulster to Argyll, 'the coastland of the Gael' (Bannerman, 1974). This community subsequently grew by absorption of the

Picts in the east and the conquest of the Britons and Angles in the south into what came to be called Scotland by the 11th century. Viking settlements in the Northern Highlands and Northern Isles from the end of the 8th century established the Norn language, which survived in Caithness, Orkney, and Shetland until the 18th century.

Under the kingship of Malcolm III, 'Ceannmòr' (1054–1096) Gaelic began to lose its pre-eminence at court and among the aristocracy to Norman French and in the Lowland area to the Anglian speech of the burghs, which were established first in eastern Scotland by David I (1124–1153). This speech was known first as Inglis and later as Scots, and it rapidly

became the predominant language of the Scottish Lowlands. However, Gaelic was maintained until the later Middle Ages in Galloway and Carrick in the southwestern Lowlands, reputedly finally ceasing in Ayrshire in the 18th century.

Linguistic Characteristics of Scottish Gaelic and its Dialects

Scottish Gaelic (Scots Gaelic) is a Celtic language, a member of the Goidelic or 'Q-Celtic' branch, closely related to Irish and Manx. It is basically a VSO language: sentences typically begin with the verb, followed by the subject and the object. Adjectives generally follow the noun. As with other Celtic languages, personal pronouns combine with prepositions and decline for person. Verbs do not generally decline for person, and there are only 10 irregular verbs.

There are two forms of the verb 'to be'. *Bi* is the basic form for straightforward statements. It conjugates fully for tense and combines with the present participle of other verbs to form continuous tenses. The emphatic form of the verb 'to be' *is* exists only in the present-future and past tenses. These two verbs, together with prepositions and prepositional pronouns, enable a vast array of idioms to be formed, which enable actions in the 'real world' to be grammatically distinguished from abstract, mental, psychic, and emotional states.

There are many English loanwords, such as *ad* (hat), *barant(as)* (warrant), *breacaist* (breakfast), *brot* (broth, soup), *comhfhurtail* (comfortable), *geata* (gate), *mionaid* (minute of time), *paidhir* (pair), *rathad* (road), *stràid* (street), and *targaid* (target). Direct borrowings from Latin include *aingeal* (angel), *airgid* (silver, money), *crois* (cross), *eaglais* (church), *Ifrinn* (Hell), *feasgair* (evening), *gineal* (offspring), *manach* (monk). Substantial contact with Norse produced *faodhail* (ford, crossing, from Norse *vadhil*), *gocaman* (lookout, from Norse *gokman*, *gauksman*), *sgalag* (lackey, from Norse *skalkr*), *sgioba* (crew, from Norse *skip*), and *uinneag* (window, from Norse *windauga*) (MacBain, 1982: 163, 200, 310, 315, 386).

Pre-aspiration in present-day Scottish Gaelic dialects occurs before the final vowels in such words as *mac* (son, pronounced as if *mahk*), *sop* (wisp, pronounced as if *sohp*), and *sloc* (pit, pronounced as if *slohk*). This is very typical of southwest Norwegian dialects and northwestern Scottish Gaelic dialects (Marstrander, in Geipel, 1971). Historically, Scottish Gaelic dialects in the northwest with its islands were very different from dialects in the east-central and eastern Highland areas. The latter are today well

nigh extinct—and all mainland dialects are moribund. Eastern dialects did not diphthongize the long 'e' in words like *meud* (measure, pronounced *meeutt*) or *beul* (mouth, pronounced *beeal*) or intrude 's' between final 'r' and 't' in words like *tart* (thirst, pronounced *tarst*), *neart* (strength, pronounced *nyarst*), etc. Similarly, the northwestern dialects intrude 't' between initial 's' and 't' as in *sruth* (stream, current, pronounced *strooh*) and *srath* (strath, wide valley, pronounced *strah*).

As with Irish, Scottish Gaelic observes the spelling convention of *caol ri caol is leathan ri leathan* (narrow to narrow and broad to broad). Where a narrow vowel 'i' or 'e' occurs before a consonant, it must be followed by a narrow vowel. Where a broad vowel 'a,' 'o,' or 'u' occurs before a vowel, it must again be followed by a broad vowel. Pronunciation of consonants is determined by the surrounding vowels; that is, the consonants thus flanked by narrow or broad vowels are regarded as correspondingly 'narrow' or 'broad' and pronounced accordingly. For example, *feasgar* (evening, afternoon) is pronounced as *fessgar*, where the first *a* is silent but helps render the consonants as 'broad.' Another example is *seillean* (bee), where the first *o* renders the initial *s* as narrow and pronounced as *sh*, and the *i* and second *e* renders the *ll* narrow and pronounced with the tongue to the front of the palate.

Gaelic uses only these letters: *a b c d e f g i l m n o p r s t u*. The letter *h* is not regarded as a regular letter. No words begin with it, except Norse-derived place names, such as *na Hearadh*, (Harris); it never stands alone in spelling, except in combination with another consonant to signify distinctive phonemes or before words beginning with vowels (e.g., *bun na h-aibhne*, the foot of the river) or to indicate aspiration or lenition in grammatical change for gender, tense, case, or word combination.

History: Medieval to Modern

By the later Middle Ages, Gaelic had retreated to the Highlands and Hebrides, which maintained some degree of independence within the Scottish state. Attempts were made by legislation in the later medieval and early modern period to establish English at first among the aristocracy and increasingly among all ranks by education acts and parish schools. The Scots Parliament passed some ten such acts between 1494–1496 and 1698. The Statutes of Iona in 1609–1610 and 1616 outlawed the Gaelic learned orders and sought to extirpate the 'Irish' language so that the "vulgar English tongue" might be universally planted (MacKinnon, 1991). The suppression of the Lordship of the Isles (1411), the Reformation (1560), the final

failure of the Jacobite cause (1746), and the end of the clan system were all in turn inimical to Gaelic.

Further setbacks for the language were brought about by the loss of life in the Napoleonic Wars, the ensuing Highland Clearances, the potato famine in the 1840s, and economic marginalization and underdevelopment that engendered large-scale migration to the Lowlands and overseas. Some mitigation resulted from legislation following the 'Crofters' Wars' in 1886, and at the end of the 19th century Gaelic was still the predominant language throughout the mainland Highlands and Hebrides.

In World War I, losses of life at sea and in the armed forces took a considerable toll on the Gaelic population, and the interwar period witnessed renewed emigration, especially from the Hebrides. The numbers of Gaelic speakers declined precipitately from 254 415 in 1891 to 58 969 in 2001. Because of internal migration from the Highlands and Islands to the Lowlands, 45% of all Gaelic speakers today reside in Lowland, in urban Scotland.

Gaelic in Present-Day Society: Cultural and Administrative Infrastructure

Although only 93 282 (1.84%) of Scotland's 5 062 011 population had any sort of oral, literate, or comprehension abilities in Gaelic in 2001 (GROS, 2001 Census), speakers and supporters have made increased demands for the use of Scottish Gaelic in education, broadcasting, and the arts, and have advocated for official recognition of the language. There was some presence of Gaelic from the earliest years of radio and on television since the mid-20th century. However, radio output of Scottish Gaelic greatly increased with BBC Radio nan Gàidheal from the mid-1980s, and a more realistic television budget from the early 1990s led to its increased use on television. Now there are demands for 24-hour daily radio provision and a dedicated digital television channel in Scots Gaelic.

Gaelic has been taught as a specific subject in some Highland and Island schools since the early 20th century, and bilingual education started in the early primary stages in Gaelic areas in the late 1950s. Although a more all-through model was introduced in the Western Isles from 1975, it has not yet really produced a satisfactorily bilingual secondary stage. Gaelic medium primary education began in 1985 after a successful Gaelic preschool initiative (*Comhairle na Sgoiltean Àraich*/CNSA from 1982) in two schools at Inverness and Glasgow. By 2004, the number of schools had grown to 60, with some 1972 pupils. In the secondary stage, 36 schools were teaching 974 fluent speakers, and 14 schools had

Gaelic medium streams with 284 pupils. A Gaelic higher education college, *Sabhal Mòr Ostaig*, was established in 1972, offering diploma courses taught through Gaelic since 1983, and a range of Gaelic medium degree courses have been offered within the developing University of the Highlands and Islands since 1998. Although these provisions have produced some small growth in the numbers of young people with Gaelic abilities, these efforts have been clearly insufficient in stabilizing Gaelic numbers overall or in reversing language shift.

Over the past 30 years, the Gaelic cultural scene has been enriched by the growth of theater and television production companies. They have been greatly assisted by funding from the Scottish Arts Council, the Gaelic arts agency *Proiseact nan Ealan* (from 1987), the Gaelic Television (now, Broadcasting) Fund, and Gaelic Media Services and its predecessors from 1992. These have drawn upon a wealth of traditional culture, including folk songs and vernacular verse. These genres have grown in Gaeldom out of the suppression of the bardic schools in the early 17th century and are still viable today. The more formal verse of the bardic period and later is well represented in current publications, as are more recent genres, such as plays and the novel. These have been assisted since 1968 by the Gaelic Books Council. The Gaelic cultural organization, *An Comunn Gàidhealach* (The Highland Association), was founded in 1891 and has run a national cultural festival, the Royal National Mod, from 1892. More recently, a local Gaelic folk festival *Fèis Bharraigh* (1980/81) has developed into a national organization, *Fèisean nan Gàidheal*, which organizes local and national Gaelic cultural festivals.

The end of the 20th century witnessed much enhancement of cultural infrastructure for Gaelic, with organizations for Gaelic learners (CLI from 1982), pre-school education (CNSA from 1982), Gaelic parents (*Comunn nam Parant*, from 1983), and language development (*Comunn na Gàidhlig* [CNAG] from 1984). These bodies have called for further resources for the language, have commissioned studies and reports, and have advocated for its official recognition and status. The 1997 Labor Government appointed a Minister for Gaelic and set up the Milne Gaelic Broadcasting Taskforce from 1997–2000. The Scottish Executive set up two more commissions: the Macpherson Taskforce on Public Funding of Gaelic (1999–2000), and the Ministerial Advisory Group on Gaelic (MAGOG; 2000–2002). These have resulted in improved provisions for Gaelic education, calls to improve Gaelic media provision, the establishment of a Gaelic Language Board, *Bord na Gàidhlig*, and the introduction of a Gaelic Language Bill in 2003. With these developments, there has been

an emphasis on research, and the application of such ideas as language planning, secure status, and reversing language shift (MacKinnon, 2004). New ideas and policy objectives have emerged in the new millennium. The future of Gaelic as a continuing language of home and community very much depends upon their outcome.

Bibliography

Bannerman J (1974). *Studies in the history of Dalriada*. Edinburgh: Scottish Academic Press.

Geipel J (1971). *The Viking legacy – the Scandinavian influence on the English and Gaelic languages*. Newton Abbot: David & Charles.

General Register, Office (Scotland) (2001). *Census Gaelic reports 1881–2001, census Scotland 2001*. Edinburgh: Register House.

MacBain A (1982). *An etymological dictionary of the Gaelic language*. Glasgow: Gairm Publications.

MacKinnon K (1991). *Gaelic: a past and future prospect*. Edinburgh: Saltire Society.

MacKinnon K (2004). 'Reversing language shift: Celtic languages today – any evidence?' *Journal of Celtic Linguistics* 8, 109–132.

Office for National Statistics (England, Wales) (2001). *Census 2001*. Titchfield: Hants.

Semitic Languages

J Hackett, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The Semitic languages are part of the Afroasiatic family. In the ancient world, Semitic languages were spoken from the western Mediterranean in the west to Iraq in the east, and from Ethiopia north to Anatolia. Many Semitic languages are still spoken today. Arabic is by far the most common; some dialect of Arabic is spoken by some 200 million speakers, from Morocco to Tajikistan, and it is also used, in its Classical and Modern Standard forms, for religious and other formal purposes. Modern Standard Arabic is the official or national language of countries throughout the Middle East and northern and north-eastern Africa, and Classical Arabic has been and still is used for religious purposes all over the world, following the spread of Islam. Modern Ethiopic languages like Amharic, Tigrinya, and Tigre are spoken by 25 million people in Ethiopia and Eritrea. Modern Hebrew is the language of 5 million inhabitants of Israel. In Yemen and Oman, Modern South Arabian languages like Mehri, Jibbali, and Soqotri have around 60 000 speakers. And Aramaic dialects continue as the languages of a few hundred thousand speakers who have left the Middle East in recent years and spread far and wide.

Our earliest attestations of a Semitic language occur in Sumerian texts of the first half of the 3rd millennium. Sumerian is not a Semitic language, but within these early texts we can recognize Akkadian names and Akkadian loanwords into Sumerian. By

around 2500 B.C., entire texts written in Akkadian start to appear.

Proto-Semitic is a scholarly reconstruction that suggests a common source to which all the known languages of the Semitic language family can be traced. Although the Semitic languages are thought to descend from some common ancestor, Proto-Semitic is not that ancestor; the term is not meant to represent a language that was ever spoken. It is instead the most economical reconstruction from which the known languages could have developed, through well-established phonological, morphological, and syntactical rules. We postulate a tense-mood-aspect verbal system with a perfective conjugation (**yaqtul*) and an imperfective conjugation (**yaqattal*). There was also a verbal adjective (**qatil* + enclitic pronouns). Another characteristic of the Semitic verbal system is a set of derived verbal stems, derived from the basic stem (the G, for *Grundstamm*), including one that doubles the middle root consonant (the D stem), a causative stem with some sort of causative affix (the C stem), a passive stem with infix or prefixed *-n-* (the N stem), a medio-passive stem with infix or prefixed *-t-* (the various *t*-stems), and possibly others. Nominals in Semitic have case endings (nominative, genitive, accusative in the singular; nominative and oblique in the plural), two genders, and three numbers (singular, dual, plural), plus both bound and unbound states of a noun.

These languages typically form words around triconsonantal 'roots.' Although there are words that cannot be related to a verbal root at all, and others that appear to have developed from fewer than three consonants, the vast majority of lexical items are formed by patterns of vowels and affixes interdigitated into and

around the three consonants that carry the meaning of a given root. Thirty consonant phonemes are reconstructed for Proto-Semitic: bilabials p, b, m, and w; interdental θ, θ', and ð; dental/alveolars t, t', d, 's, 's', 'd, s, 't, 't', l, r, and n; palatal y; velars k, k', g, x, x, and y; pharyngeal h and 'h; and glottal ' and h. We reconstruct three short and three long vowels: a, i, u, ā, ī, ū.

Verbal sentences in Semitic are typically V-S-O, and adjectives follow the nouns they modify. There is a genitive chain, called a construct chain, that consists of two or more nouns, the first of which can be in any case, but the remaining members are genitive. For the most part, this chain describes the 'of' relation: for instance, 'the king of the land' would be 'king' + case ending followed by 'the land' + genitive case ending. Adjectives and the nouns they modify must agree in gender and number, except that the numerals from three to ten have the odd feature that the feminine-looking numeral modifies the masculine noun, and the masculine-looking numeral modifies the feminine noun. Some of these features can be followed through the innovations that define the various branches of Semitic.

East and West Semitic

The Semitic languages are a fairly close-knit family. Semitic divides into two major groupings, East Semitic and West Semitic.

East Semitic

East Semitic is made up of Akkadian and Eblaite.

Akkadian Akkadian is the language of ancient Mesopotamia (ancient Iraq), and so far hundreds of thousands of Akkadian texts have come to light. It is written in left-to-right cuneiform, inscribed with a stylus on clay tablets. Akkadian cuneiform uses logograms and syllabic signs, with vC, Cv, and CvC syllables. Akkadian can be broken down into several dialects. In the 3rd millennium, we use the term Old Akkadian for a number of dialects used to write royal inscriptions, letters, ritual texts, administrative texts, and literary texts. Beginning in the early 2nd millennium, the umbrella term Akkadian is replaced by Assyrian and Babylonian. Old Assyrian is the language of some 15 000 early 2nd-millennium documents, mostly written by or for merchants residing in Anatolia. The name Old Babylonian covers the many thousands of texts from the first half of the 2nd millennium, especially from the First Dynasty of Babylon, of which Hammurabi is the most famous king. These texts are letters, omen texts, literary texts, administrative texts, and laws, including the famous law code of Hammurabi.

Both Middle Assyrian and Middle Babylonian, from the second half of the 2nd millennium, are less fully documented, but with the same array of genres. For much of the 2nd millennium, Akkadian was a lingua franca for the entire Near East, and Peripheral Akkadian texts are found from Egypt (especially the el-Amarna archive of letters from Palestinian governors) to Syria (especially the administrative texts from ancient Ugarit, modern Ras Shamra) to Anatolia (the archive from ancient Hattusa, modern Boghazköy).

Neo-Babylonian texts survive from the first half of the 1st millennium, largely letters and administrative texts, and Late Babylonian continues in use until the 1st century A.D. Neo-Assyrian, the language of the Neo-Assyrian Empire, stretches from the early 1st millennium until the fall of the empire in the late 7th century B.C. These documents are far more numerous than the Babylonian texts of the 1st millennium. Finally, Standard Babylonian refers to the archaizing written language in use in the first half of the 1st millennium, in which both Babylonians and Assyrians recorded religious and literary texts, royal inscriptions, and other formal texts.

Eblaite A cache of texts was uncovered beginning in the 1970s at Tell Mardikh in Syria, the ancient city of Ebla. These texts, which date to the 24th or 23rd century B.C., are largely in Sumerian, but bilingual lexical lists and some other texts display another language that is not Sumerian and not Akkadian, but seems to be closely related to Akkadian. It is this language that is dubbed Eblaite.

West Semitic

The West Semitic languages are separated from the East Semitic by an innovation that can be seen in all West Semitic languages: the development of an original verbal adjective **qatil* + case ending into an active, perfective suffix conjugation, **qatala*. This new perfective conjugation replaces the Common Semitic perfective/volitive form **yaqtul*, which continued in use in West Semitic, especially as a volitive, but as a past tense form only in restricted environments. West Semitic itself can be divided into the Ethiopian languages, the Modern South Arabian languages, and the Central Semitic languages.

Ethiopian Classical Ethiopic (Ge'ez) is attested beginning in the 4th century A.D. as the language of ancient Aksum and probably went out of use as a spoken language in the 10th century, with the demise of the Aksumite Empire. It continued, however, as the language of the Ethiopian church and as a general

literary language until recently. Closely related to Ge'ez are the Northern Ethiopian languages Tigrinya and Tigre, spoken in Eritrea. Amharic, the official language of Ethiopia, is the best known of the Southern Ethiopian languages. It is generally thought that the Ethiopian languages came to east Africa from the southwest Arabian peninsula, probably no earlier than the 1st millennium B.C. The earliest Ge'ez was written in the alphabet used for Old South Arabian inscriptions; this alphabet later developed into a distinctive Ethiopic syllabary with the addition of vowel marks.

Ethiopian languages use the innovated **qatala* perfective form, but are differentiated from the Central Semitic by their retention of the old Common Semitic imperfective form **yaqattal*, with doubled middle radical, which they share with East Semitic.

Modern South Arabian The Modern South Arabian languages (particularly Mehri, but also Jibbali, Hobyót, Harsusi, Soqotri, and several smaller groups), are spoken by around 60 000 people (and dwindling) in Yemen and Oman. They have only recently been written down and so have no literary history; there is also no script associated with them, since we know them from transcriptions into the Latin or Arabic alphabet. Like the Ethiopian languages, Modern South Arabian languages have both the innovated **qatala* perfective and the retained **yaqattal* imperfective. These languages are not the descendants of the Epigraphic or Old South Arabian inscriptions (see below).

Central Semitic The Central Semitic languages break off from the rest of West Semitic with the innovation of a new imperfective form **yaqtulu*, probably a development from an old subjunctive form such as we see in Akkadian **yaqtul-u* used in subordinate clauses. The Central Semitic languages also exhibit the innovated **qatala* perfective that defines all West Semitic languages. Volitive **yaqtul* remains but can be confused with imperfective **yaqtulu* in languages without final short vowels; preterite **yaqtul* is also used in certain restricted environments such as the *waw*-consecutive preterite *wayyiqtol* forms in Classical Hebrew. The Central Semitic branch is divided into Old South Arabian, North Arabian, and Syro-Palestinian or Northwest Semitic.

Old South Arabian The Old South Arabian (or Ancient South Arabian) inscriptions (or Sayhadic) date from the 8th century B.C. to the 6th century A.D. This umbrella term includes several dialects, the best attested of which is Sabaean (or Sabaic); inscriptions

are also found in Hadramitic, Minaean, and Qatabanian dialects, but in much smaller numbers. There are at present more than 10 000 stone inscriptions in Sabaean, and a recent find of many inscriptions on short wooden sticks in a cursive script, probably from the 2nd or 3rd century A.D., has greatly increased our knowledge of the grammar of this language.

Recent work has shown that the language of the Old South Arabian inscriptions should be classified along with Central Semitic rather than the traditional South Semitic classification. I-*w* verbs in these inscriptions never have *w* in the prefix conjugation, and since the **yaqattal* form would show the *w*, the prefix conjugation must be not **yaqattal*, but rather the Central Semitic **yaqtul*.

The earliest Sabaean inscriptions are mainly written boustrophedon; otherwise, writing is from right to left. The majority of the stone inscriptions fall into the following categories: graffiti, mostly personal names; dedicatory inscriptions; building inscriptions; reports of military campaigns; legal documents; funerary inscriptions. The inscriptions on wood at least partly concern legal and economic matters, sometimes written in the form of a letter.

The South Semitic alphabet, which broke off from the early Canaanite linear alphabet around the 13th century B.C., represented only consonants. The Old South Arabian inscriptions maintain this system of writing, with rare indications of vowels.

North Arabian The North Arabian languages of the Central Semitic branch are divided into two groups: Arabic, including Old Arabic, Classical, and modern dialects; and Old North Arabian (or Ancient North Arabian).

Old North Arabian includes inscriptions in pre-Islamic dialects, dating from the 8th century B.C. to the 4th century A.D.: Oasis North Arabian (Taymanite, Dedanite, Dumaitic, and Dispersed Oasis North Arabian); Safaitic; Thamudic; and Himaic; plus what is generally called Hasaitic – a dialect seen in a few dozen mostly funerary inscriptions from northeastern Arabia, near the Persian Gulf, probably dating from the second half of the 1st millennium B.C. Old North Arabian inscriptions are largely graffiti, and so the handwriting is mostly informal, not highly trained; therefore, the dating and even ancestor of some of the individual scripts are not clear. For the most part, however, Old North Arabian inscriptions are written in a script closely related to that of the Old South Arabian inscriptions. The number of graffiti and their dispersal patterns are astounding in their implications for literacy, both in the towns and among the nomads, since thousands of the graffiti so far published were written on rocks in the desert areas

that stretch from Syria to northern Arabia. The writing of the Oasis texts was usually right to left, and sometimes boustrophedon, while the nomads' writing on rocks goes in every possible direction, including a spiral, across the surface of the rocks on which they were written.

The Oasis dialects are those used in the oases associated with the routes of the spice and aromatics trade (especially frankincense). Approximately 400 short inscriptions found in and around the oasis of Tayma have been published so far, in a dialect and script that are somewhat different from the rest of Old North Arabian. They probably date to the 6th or 5th century B.C. Two inscriptions have recently been published that mention "Nabonidus, king of Babylon"; Nabonidus in fact spent ten years of his reign in Tayma in the mid-6th century B.C. Dedanite is the dialect of hundreds of graffiti found in the vicinity of the oasis of Dedan. It is the only one of these dialects used for monumental inscriptions as well, and so far hundreds of mostly dedicatory monumental inscriptions have been found. Dedanite inscriptions probably date to the second half of the 1st millennium B.C. The Dumaitic dialect is represented by only three texts so far, found in or near Sakaka in northern Arabia. There is very little evidence for dating, but they may also be from the second half of the 1st millennium B.C. The Dispersed Oasis North Arabian are those texts that are clearly related to the Oasis North Arabian but are found outside Arabia, carried north by traders and found mostly in what was Mesopotamia.

Of the Safaitic, Thamudic, and Hismaic inscriptions, the Safaitic are the most numerous, with some 15 000 graffiti from the 1st century B.C. to the 4th century A.D.; written by nomads, they are found as far north as Damascus and east to the Euphrates, as well as south to northern Arabia. Hismaic inscriptions are written in the language of the Hisma desert nomads, and they stretch from northwestern Arabia to central and northern Jordan. Approximately 1000 Thamudic graffiti date from the 6th century B.C. to the 3rd century A.D. Thamudic includes all the non-Oasis Old North Arabian inscriptions that are not otherwise classified.

The prominent difference between Old North Arabian and Arabic is the definite article in each: *ʾal-* in Arabic, but *h-* in Old North Arabian.

Classical Arabic has made its way around the world as the literary language of Islam, but a dialect or dialects similar to Classical Arabic were known already in 4th-century A.D. inscriptions (hence the name Old Arabic). Classical Arabic has elements of spoken pre-Islamic Arabic, but was shaped by the Qurʾan (7th century A.D.), which reflects also the spoken dialect of the Hejaz region of central and

west Arabia, in particular, the dialect of Mecca which was Muhammad's dialect. Arabic script was adapted from the script of the Nabateans. It is written from right to left and was originally consonantal, with indication of long vowels, and has developed diacritical marks to indicate short vowels and other features. The 8th- and 9th-century Arab grammarians standardized the language, and it has changed very little since that time.

Modern Standard Arabic, which is the Arabic of newspapers, radio, television, and international communication, is still Classical Arabic at its base, with vocabulary updated as necessary. A development from Classical Arabic is Middle Arabic, which is meant to be literary Arabic, but deviates from it in ways that betray the authors' dialects; examples of Middle Arabic are Judeo-Arabic, Medieval Christian Arabic, and Spanish Arabic. Finally, there are the modern vernacular dialects of Arabic that have evolved over the centuries over the large territory in which Arabic is spoken. Many, in fact, are no longer mutually comprehensible, especially at the outer ends of that territory, so that someone speaking in the vernacular dialect of Morocco and someone speaking in the vernacular dialect of Iraq would have to move closer and closer to the classical language until they came to a point where each could comfortably understand the other. Arabic is one of the most commonly spoken languages in the world, with approximately 200 million speakers.

Northwest Semitic The Northwest Semitic (or Syro-Palestinian) languages can be divided into at least four subcategories: Ugaritic, Canaanite, Aramaic, and other. These languages all share the innovations of West and Central Semitic, plus two more: the change of initial **w* to **y* in virtually all environments except for the word 'and,' which remains proclitic **wa-*; and the plural pattern that adds *-a-* insertion to the regular external plural, for nouns of the type $C_1VC_2C_3$. Thus, **sipr-* 'book'; plural **siparūma/siparīma* (Hebrew *šēper, šepārīm*).

Ugaritic is the language attested from the late 14th century to around 1200 B.C. at the ancient city of Ugarit (modern Ras Shamra) on the Syrian coast above Latakia, and at an outlying town called Ras Ibn Hani. It is known from the approximately 1500 texts written left to right in alphabetic cuneiform on clay tablets. The writing system postdates the linear alphabet known from the same general region and seems to be a clever combination of the technology of cuneiform writing on clay (like the lingua franca at the time, Akkadian) and the idea of an alphabet and the ease of writing it affords. Ugaritic is written without vowels, but there are some multilingual texts with

columns that represent the Ugaritic pronunciation written out in syllabic cuneiform, so that the vocalization of some words and some basic rules are known. Further, there are three signs for 'aleph, representing, among other things, the consonant 'aleph plus vowels a, i, and u, so words with 'aleph as a root letter often have some indication of the vocalization of that word and by extension other words of the same type (perfective verb, for instance). The corpus consists of poetic mythological texts, ritual texts, administrative texts, letters, and school texts.

Ugaritic is sometimes thought to be a Canaanite language rather than a branch on its own, but there is at least one of the defining innovations of Canaanite that Ugaritic does not participate in (see below on 'intensive' and causative stems); perhaps there were more, but they cannot be seen in the Ugaritic consonantal alphabet, even with the aids mentioned above. Canaanite describes a grouping of closely related languages: early Canaanite seen in scattered inscriptions and in the underlying dialects of the Amarna letter scribes (see below); Phoenician; Moabite; Hebrew; Ammonite; and Edomite. These languages share a change of the 'intensive' and causative conjugation perfective verb forms from **qattila* and **haqtilla* to **qittila* and **biqtilla*, and the so-called Canaanite shift of *ā* to *ō*, among others. While these languages are written consonantally, later vocalizations (of Biblical Hebrew, for instance) and the spelling out of words in contemporaneous Akkadian, Greek, and Roman documents allow us to reconstruct many of these changes where they are not obvious.

Ancient Canaan covered roughly modern southern Lebanon, Israel, and the northwestern part of Jordan. The earliest indications of Canaanite languages come from the few scattered alphabetic inscriptions beginning in the 18th century B.C., and from the 'Amarna letters,' Akkadian cuneiform clay tablets written by rulers of ancient Canaan (among others) to the pharaoh in Akhetaten, modern el-Amarna, in Egypt. Sometimes the Canaanite-speaking scribe glosses a word in his own language, using Akkadian syllabic writing just as he did for the lingua franca Akkadian in which the texts are mostly written; and this Akkadian itself is a mixed language with Akkadian vocabulary but the morphology and syntax of the local dialects of the scribes. There is enough evidence in the Amarna letters to identify the dialects of the Canaanite scribes as an early form of what later became Phoenician, Hebrew, and so on.

Phoenician is the name we give to the Canaanite language spoken in the cities of the northern coast of the Levant; modern scholars refer to the language as Phoenician after about 1200 B.C. This date marked a turning point in the region's fortunes, because it was

approximately then that the so-called Sea Peoples, refugees from Mycenaean Greece, attacked Egypt, establishing a number of cities on the southern coast of the Levant and freeing the rest of Canaan from Egyptian control. (The most famous of the Sea Peoples are the Philistines.) Phoenician spread throughout the Mediterranean through the trading empire of the Phoenicians, and later the commercial empire of North African Carthage, originally an outpost of Levantine Tyre. Phoenician is a long-lived language: the first inscriptions long enough to analyze come from 10th-century Byblos, and in its Carthaginian extension (called Punic), the language is attested until the 4th century A.D., in Latino-Punic inscriptions. Phoenician and Punic texts are dedicatory, royal, funerary, votive, and commercial, plus a number of seals and coins.

Hebrew is the best-known Canaanite language because it was used to write most of the Hebrew Bible; because of its continued use as a literary language throughout Late Antiquity, the Middle Ages, and beyond; and because of its modern incarnation as the language of the modern state of Israel. Biblical or Classical Hebrew was the language of the ancient kingdoms of Israel and Judah in at least two dialects, Northern (or Israelian) and Judahite, from around 1200 B.C. to 600 B.C., when Judah was defeated by the Neo-Babylonian Empire. It is known from texts in the Hebrew Bible and from epigraphic remains and was written in the Hebrew script (developed from the linear Canaanite alphabet) right to left on various media, including stone, potsherds, and metals; no doubt the majority of ancient Hebrew texts were written on papyrus and are now lost to us. The few epigraphs that remain from this period are commercial, dedicatory, funerary, and administrative, plus a few seals and bullae (only a few that are surely authentic; forgeries abound).

It remained the spoken language of the area of those former kingdoms on some level, but exactly how central it was to most people is not clear, since Aramaic (see below) had taken over as the lingua franca in the Near East as early as the 8th-century Neo-Assyrian Empire. Once Israel and Judah became provinces in the Assyrian, Babylonian, and Persian empires, most epigraphic evidence appeared in Aramaic, and the books of the Hebrew Bible that are written in that era, such as Daniel, Esther, and Chronicles, are written in Late Biblical (or Classical) Hebrew, a dialect or dialects that are either simply different from or developed from the Standard Biblical Hebrew of the earlier literature of the Bible. The extent of the differences is muted by the many editorial hands that have leveled the dialects in which the Hebrew Bible was written.

Middle Hebrew was the language of the several nationalistic revolts (under the Maccabees in the 2nd century B.C.; against the Romans, ending in 70 A.D.; and the Bar Kokhba insurrection ca. 135 A.D.) and it continued to be used in written texts from the 2nd century B.C. to the 5th century A.D., including the Hebrew texts from Qumran, Samaritan Hebrew, and the Mishnah. Medieval Hebrew continued the tradition of Middle and Classical Hebrew, both as a sacred and literary language, and as the language in which Jews from different countries could communicate. Modern Hebrew was revived as a spoken language in the 19th century and is still spoken today in the state of Israel.

Besides a few seals, Moabite is known from one reasonably long inscription, written on behalf of King Mesha to commemorate his military triumphs. It dates from the second half of the 9th century and is very much like Hebrew and Phoenician. The area of ancient Moab was in the southern part of modern Jordan. Ammonite and Edomite are known from only a very few inscriptions and seals, but as far as they can be analyzed, they, too, are very similar to the better-attested Hebrew and Phoenician. Ancient Ammonite was spoken in the central highlands of what is now Jordan, and Edomite was the language of the far south of Jordan.

Aramaic is distinguished from the rest of Northwest Semitic by its use of *-na* as the first-person plural suffix in all environments; by the change from vocalic *n* to *r* in the words for 'son,' 'daughter,' and 'two'; by its loss of the N-stem; and by the development of a new causative-reflexive stem **hittaqtal*, which replaces general Central Semitic **(v)staqtala*. Aramaic is divided into Old Aramaic, Official or Imperial Aramaic, Middle Aramaic, Late Aramaic, and Modern Aramaic.

Old Aramaic describes inscriptions from a number of closely-related dialects that date from the mid-9th century to the 6th century B.C. They are royal inscriptions, for the most part, but there is also a long treaty text, funerary inscriptions, religious texts, and a few seals, as well as other scattered pieces. Most are carved into stone in a 22-letter alphabet borrowed from the Phoenicians, written right to left, that is mostly consonantal, but even early on with indications of some long vowels.

Official or Imperial Aramaic was the lingua franca of the Persian Empire (from the 6th to the 4th century B.C.), and as such it spread west to Egypt and east as far as Pakistan. It survives in large numbers of papyri, especially from Egypt, where they have been protected by the dryness for millennia. The Aramaic of the biblical book of Ezra is Imperial Aramaic.

Known from the 3rd century B.C. to the 2nd century A.D., Middle Aramaic describes a large number of

inscriptions written in a variety of dialects, such as Nabatean, Palmyrene, Hatran, and Old Syriac. Standard Literary Aramaic developed in Palestine and is the Aramaic of the biblical book of Daniel, of Targums Onqelos and Jonathan, of the Dead Sea Scrolls, the Bar Kochba letters, and quotations in the Mishnah and New Testament.

Late Aramaic begins in the 3rd century A.D. and includes Late Western Aramaic, Late Eastern Aramaic, and Literary Syriac. Late Western Aramaic is the language of the Palestinian Talmud, the Midrashim, and the Targums. It also includes Christian Palestinian Aramaic and Samaritan Aramaic. Late Eastern Aramaic is the language of the Babylonian Talmud. It also includes Mandaic and the language of a large number of incantation bowls from the 4th to the 7th century A.D. There is, finally, a huge Christian literature written in Literary Syriac from the 4th to the 13th century A.D. After the rise of Islam in the 7th century, Syriac was used less and less as a spoken language, but continued as the literary language of the Syrian Orthodox Church.

Modern Aramaic, or Neo-Aramaic, is still spoken by a few hundred thousand people from communities formerly situated in Iran, Iraq, and Syria for the most part; in recent decades, most of the speakers have left the Middle East and emigrated with their languages to the United States, Sweden, Germany, the Netherlands, and Australia, among others. Neo-Aramaic speakers are Christian, Muslim, Jewish, and Mandaean. The four main branches of Neo-Aramaic are Western (spoken in three villages near Damascus); Central (Turoyo and Mlahso, spoken in southeastern Turkey); Eastern (or Neo-Syriac, no relation to classical Syriac; spoken originally in Kurdistan); and Neo-Mandaic (spoken originally in western Iran).

There are some Northwest Semitic inscriptions that are difficult to identify as Ugaritic, Canaanite, or Aramaic. The dialect of the 8th-century B.C. prophetic inscription found at Tell Deir Alla in Jordan has been a matter of much controversy because it seems to combine features of Aramaic dialects and of Canaanite dialects. The same is true to a lesser degree of the dialect of two 8th-century royal inscriptions from Zincirli in Turkey. It has been suggested that these inscriptions represent other forms of Northwest Semitic that did not hivel off of the main branch at quite the same time as the three major languages.

Bibliography

Beeston A F L (1992). 'Languages: pre-Islamic South Arabian.' In Freedman D N (ed.) *Anchor Bible dictionary*, vol. 4. New York: Doubleday. 223–226.

- Bergsträsser G (1928). *Einführung in die semitischen Sprachen*. Munich: M. Hueber. Daniels P T (trans.) (1983). *Introduction to the Semitic languages*. Winona Lake, IN: Eisenbrauns.
- Daniels P T & Bright W (eds.) (1996). *The world's writing systems*. New York: Oxford University Press.
- Garr W R (1985). *Dialect geography of Syria-Palestine, 1000–586 B.C.E.* Philadelphia: University of Pennsylvania.
- Grimes B (2003). 'Language list' [included in the article on Semitic languages]. In Frawley W J (ed.) *International encyclopedia of linguistics*, 2nd edn., vol. 4. Oxford: Oxford University Press. 33–40.
- Hackett J A (2002). 'The study of partially documented languages.' In Izre'el S (ed.), 57–75.
- Hackett J A (2004). 'Phoenician and Punic.' In Woodard R (ed.), 365–385.
- Hetzron R (1997). *The Semitic languages*. London and New York: Routledge.
- Hetzron R (2003). 'Semitic languages.' In Frawley W J (ed.) *International encyclopedia of linguistics*, 2nd edn., vol. 4. Oxford: Oxford University Press. 29–33.
- Huehnergard J (1992). 'Languages: introductory survey.' In Freedman D N (ed.) *Anchor Bible dictionary*, vol. 4. New York: Doubleday. 155–170.
- Huehnergard J (1995). 'Semitic languages.' In Sasson J M (ed.) *Civilizations of the ancient Near East*, vol. 4. New York: Scribner's. 2117–2134.
- Huehnergard J (2003). 'Comparative Semitic linguistics.' In Izre'el S (ed.), 213–254.
- Huehnergard J & Woods C (2004). 'Akkadian and Eblaite.' In Woodard R (ed.), 218–280.
- Izre'el S (ed.) (2003). *Semitic linguistics: the state of the art at the turn of the twenty-first century*. Israel Oriental Studies 20. Winona Lake, IN: Eisenbrauns.
- Johnstone T M (1975). 'The modern South Arabian languages.' *Afroasiatic Linguistics* 1, 93–121.
- Kaltner J & McKenzie S L (eds.) (2002). *Beyond Babel: a handbook for Biblical Hebrew and related languages*. Atlanta: Society of Biblical Literature Press.
- Leslau W (1970). 'Ethiopic and South Arabian.' In Sebeok T (ed.), 467–527.
- Macdonald M C A (2004). 'Ancient North Arabian.' In Woodard R (ed.), 488–533.
- Moran W L (1961). 'The Hebrew language in its Northwest Semitic background.' In Wright G E (ed.) *The Bible and the ancient Near East: essays in honor of W F Albright*. New York: Doubleday.
- Nebes N & Stein P (2004). 'Ancient South Arabian.' In Woodard R (ed.), 454–487.
- Pardee D (2004). 'Ugaritic.' In Woodard R (ed.), 288–318.
- Sebeok T (ed.) (1970). *Current trends in linguistics*, vol. 6. The Hague/Paris: Mouton.
- Ullendorff E (1970). 'Comparative Semitics.' In Sebeok T (ed.), 261–273.
- Voigt R (1987). 'The classification of Central Semitic.' *Journal of Semitic Studies* 32, 1–21.
- Woodard R (ed.) (2004). *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press.

Serbian-Croatian-Bosnian Linguistic Complex

E C Hawkesworth, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Bosnian

The official name of the language spoken in Bosnia Herzegovina is Bosnian. The status of the language in reality, however, is more complex, as may be seen in the language law adopted in 1993: "In the Republic of Bosnia and Herzegovina, the Ijekavian standard literary language of the three constitutive nations is officially used, designated by one of the terms: Bosnian, Serbian, Croatian. Both alphabets, Latin and Cyrillic, are equal." The law reflects the fact that the territory is inhabited by three national groups: Bosniaks (South Slav Muslims, the majority population of the Bosnian/Croatian Federation), Croats (Catholics, the majority population of the territory of Herzegovina,

the southwestern area of the Bosnian/Croatian Federation); and Serbs (Orthodox, the majority population of the other Bosnian entity, known as *Republika srpska*). In practice, it is normally only Bosniaks (and those committed to the survival of Bosnia Herzegovina as a unified country), who refer to their language as 'Bosnian.' And it is logical enough that Croats should speak Croatian and Serbs Serbian now that there is no longer an all-inclusive Serbo-Croatian umbrella. The debate as to whether or not a distinct Bosnian language exists continues. At the time of writing, the standard language used by the official authorities in Sarajevo and other parts of the Federation may be described as distinct from the standard languages in Serbia and Croatia, but the process of standardization, through dictionaries, grammars, and scholarly studies, has yet to be completed. For the time being, it cannot be said that Bosnian has quite the same status as Croatian in terms of its recognition as a specific standard.

Croatian

Croatian is the official name of the language spoken in the territory of Croatia. To a considerable extent, the political tensions that ultimately led to the collapse of Yugoslavia were first reflected in issues related to language. Long before it was possible for ideas of political separation to be contemplated, in 1967, Croatian linguists published a Declaration on the Name and Position of the Croatian Literary Language, calling for official recognition of Croatian as a separate language. From the outset, however, it was clear that the Declaration had more to do with cultural and sociopolitical aspirations than linguistics. From 1971, nationalist policies in Croatian became steadily more entrenched, leading eventually to the secession of Croatia (and Slovenia) from the common state of Yugoslavia in 1992. As language and statehood have been inextricably linked since the rise of nation states in Europe, it was understandable that nationalist politics should place particular emphasis on separating the Croatian element of the Serbo-Croatian language as far as possible from its Serbian counterpart. To this end, archaic words were reintroduced, neologisms forged and various 'differential' dictionaries published in an effort to raise the consciousness of individual Croats to the special nature of their language and to purify the Croatian language of Serbianisms. Apart from lexical items and favoring two characteristic syntactic differences (the infinitive in Croatian for dependent verbs, as opposed to *da* + present tense in Serbian; verb + interrogative particle *li* for questions as opposed to *da li* + verb in Serbian), particular emphasis has been placed on differences in word formation. At the height of the nationalist era, in the extreme circumstances of war and later, as Croatia consolidated its position as an independent state, linguists were particularly active. Some of the results of this frenzy were inevitably artificial and at times entertaining. This phase of heightened self-consciousness has now passed, with the recognition that Croatian has been widely accepted as a separate standard at an official level. Speakers may now be left to express themselves naturally and the language to develop in a more organic manner.

Serbian

Serbian is the official name of the language spoken in the territory of Serbia and Montenegro. Unlike the other components of the Serbian-Croatian-Bosnian linguistic complex, Serbian, as the standard language of the Serbs and Montenegrins, has not changed

essentially from its earlier incarnation as Serbo-Croat. It was the Croats who opted to remove their language from the dual name, and set about making their standard as distinct as possible from standard Serbian (see 'Croatian'), while the Serbs had only to stand still. The process of the disintegration of standard Serbo-Croatian may thus be described as 'asymmetrical and asynchronous' (Ljubomir Popovic, 'From Serbian to Serbo-Croatian to Serbian,' in Bugarski and Hawkesworth, 2004). In response, a Serbian Language Standardization Committee was set up to describe the current situation and Serbian has now been officially recognized as a separate language within Slavonic studies.

Serbian-Croatian-Bosnian Linguistic Complex

The language formerly known as Serbo-Croat belongs, with Bulgarian, Slovene, and Macedonian, to the South Slav branch of the Slavonic language family. The first written records are 11th-century inscriptions in stone in both the Glagolitic and related Cyrillic scripts. The cultural division between the two variants reflects their history: the western Latin-script culture of Croatia, in the orbit of the Catholic Church and later the Hapsburg Monarchy; and the eastern, Cyrillic, Byzantine, Orthodox culture of Serbia.

The dia-system linguistic complex is the most heterogeneous Slavonic dia-system, with an exceptionally large variety of dialects, some with six or seven cases, some with four, and a great variety of verbal tenses. At the same time, these dialects have a striking degree of connectedness, containing characteristic features, which distinguish the complex from all other Slavonic languages. One of these is its archaic prosodic system, in which stress position, vocalic quantity (length/shortness) and tone (rising/falling) are marked. The traditional accents are long falling: *noć*; short falling: *kuća*; long rising: *réka*; short rising: *ostati*. There are not many minimal pairs. Examples would be *grad* 'hail' and *grad* 'town'; *pas* 'dog' and *pas* 'belt, waist; pass'; and the sentence *Sam sam* 'I [masc.] am alone'.

In terms of morphology, the structure has remained complex, although one feature of Old Slavonic – the dual – has disappeared from the declensions and conjugations of all dialects in the complex. Case and verbal endings and accent shifts are the main morphological categories.

Word order is free, with the exception of strict rules governing the position of enclitics. These are verbal and pronominal short forms and the interrogative and reflexive particles.

Orthography has experienced the systematization of the Serbo-Croatian vernacular, which was carried out in the mid-19th century on phonetic principles, with one letter corresponding to one sound, making its orthography one of the most consistent in Europe. There is exact correspondence between the two scripts so that transliteration from one to the other is straightforward. There are three symbols unique to the language: **ć, ħ; đ ĥ; dž ȣ**. For example:

Ijekavian variant (characterized by the rendering of Old Slavonic *jat* as *je* or *ije*, and spoken in Croatia, Bosnia, and Montenegro):

Od dviju sjevernih skupina, tj. istočne i zapadne, južna se razlikuje nizom osobina.

Ekavian variant (characterized by the rendering of Old Slavonic *jat* as *e*, and spoken in most of Serbia, which can equally well be written in the Latin script):

Од двеју северних скупина, тј. истоčne и западне, јужна се разликује низом особина.

'The Southern (Slavonic) group is distinguished from the two Northern groups, i.e., the Eastern and Western, by a series of features.'

Extensive bibliographies, as well as detailed studies, on the language situation of former Yugoslavia may be found in Bugarski *et al.*, 1992; the current situation is covered in the sequel: *Language in the former Yugoslav lands*, Slavica, 2004.

Serbo-Croat

The linguistic unity of the majority of the Southern Slav population of the Hapsburg and Ottoman lands that were to become Yugoslavia after the collapse of these empires in the First World War was first acknowledged in the joint Literary Agreement of 1850. The name 'Serbo-Croat' was officially adopted with the formation of the Kingdom of the Serbs, Croats, and Slovenes (known as Yugoslavia from 1928). It was never a straightforward phenomenon, however, as can be seen in the description adopted by many scholars: a polycentric standard language. The language could be officially termed 'Serbo-Croat,' 'Croato-Serbian,' 'Serbian and Croatian,' 'Croatian and Serbian,' 'Serbian or Croatian,' 'Croatian or Serbian.' In practice, from the end of the 1960s,

most people in Croatia and Serbia referred to their language as 'Croatian' or 'Serbian,' respectively, simply for convenience, without this label implying any separatist tendencies. This situation lasted until the collapse of Yugoslavia in the wars of 1991–1995. Since the establishment of the independent states of Bosnia Herzegovina, Croatia, and Serbia and Montenegro (still officially known as Yugoslavia until 2003), the term 'Serbo-Croat' no longer has any official validity in sociopolitical terms. The language spoken in these countries is now officially known as Bosnian, Croatian, and Serbian, respectively. In linguistic terms, the standard language remains essentially the same, but the sociopolitical reality is that it no longer has a single name. When native speakers wish to refer to the language in its broader sense, beyond the borders of their own homeland, they tend to say 'naš jezik' or 'naški (our language). For the purposes of the War Crimes Tribunal in The Hague, it is known as BCS. University departments in Europe where it is taught refer to it variously as Bosnian/Croatian/Serbian (Austria, Norway); Serbo-Croatian (Denmark); Serbo-Croat (France); South Slavic (Finland); Serbian/Croatian/Bosnian (Sweden); Serbian and Croatian (UK). In the absence of an entirely satisfactory solution, in this volume the term 'Serbian-Croatian-Bosnian linguistic complex' has been adopted as a clumsy but accurate description.

Bibliography

- Hawkesworth E C (1984). *Ivo Andric: bridge between East and West*. London: Athlone Press.
- Hawkesworth E C (1986). *Colloquial Serbo-Croat*. London: Routledge and Kegan Paul. Revised and published as *Colloquial Croatian and Serbian*. (1998). Reissued as *Colloquial Croatian* and published with *Colloquial Serbian*. (2005).
- Hawkesworth E C & Bugarski R (eds.) (1992). *Language planning in Yugoslavia*. Columbus: Slavica.
- Hawkesworth E C (2000). *Voices in the shadows. Women's writing in Serbia and Bosnia*. CEU Press.
- Hawkesworth E C (ed.) (2001). *A history of Central European women's writing*. Palgrave.
- Hawkesworth E C & Bugarski R (eds.) (2004). *Language in the former Yugoslav lands*. Slavica.

Shona

D Odden, The Ohio State University, Columbus, OH, USA

© 2006 Elsevier Ltd. All rights reserved.

Shona is a member of the Bantu language family, S10 in the Guthrie classification, having roughly 7 000 000 speakers in Botswana, Malawi, Mozambique, and Zambia, with the majority of speakers in Zimbabwe. Shona is the dominant African language of Zimbabwe; as one of the major languages of southern Africa, the speaker population is comparable to that of Zulu and Xhosa. Shona is also a literary language, with a considerable literature having developed since the 1950s. The main dialects of Shona are Karanga, Korekore, Manyika, and Zezuru; Zezuru forms the basis of the standard language. Ndau and Kalanga are closely related and might be considered to be highly divergent dialects or closely related but separate languages: in this article, they will not be treated as Shona dialects. The main linguistic reference works for Shona are the reference grammars of Fortune (1955, 1980) an extensive dictionary indicating tone and dialectal properties (Hannan, 1984), and *Duramazwi Guru reChiShona*, which is a major Shona monolingual dictionary.

With 5 vowels and minimally 32 uncontroversial consonants, Shona has somewhat more than the usual complement of consonants for Bantu languages. The consonantal system includes a contrast between the labiodental fricative [v], spelled *vh*, and a bilabial semi-approximant [β], spelled *v*, as well as a labiodental flap [ṽ] found in some ideophones and reported in a few nouns, as well as a contrast [b, d] vs. [ɓ, ɗ]. The most well-known phonetic oddity of Shona is a set of ‘whistling fricatives,’ a set of retracted alveolar fricatives and affricates articulated between [s] and [ʃ] (with which they contrast), produced with a degree of lip protrusion but not full rounding: these sounds, for which the International Phonetic Alphabet (IPA) lacks symbols, are spelled *sv*, *zv*, *tsv*, *dzv*. In some dialects, especially Manyika, the labial constriction is so extreme that a stop closure is formed. In addition, the language has a series of prenasalized consonants of nasal plus consonant sequences [mb], etc., and morphophonemic consonant plus *w* sequences pronounced with various degrees of velarization according to dialect, so that /tw/ may be phonetically [txw] or [tkw] with unrounding after labials, and /pw/ may be [px] or [pk].

Shona is a tone language, but extensive tonal data is available only for the Karanga and Zezuru dialects (see Fivaz (1970) and Fortune (1980) for

tone-marked paradigms of Zezuru, Myers (1987) for analysis, and Odden (1981) for Karanga). Differences in the tonal morphophonemics of different Shona dialects are significant, and are comparable to tonal differences across Makonde dialects or the so-called Luhya languages. One characteristic of Shona tonology is a panoply of dissimilative H-tone lowering and rightward spreading processes, which are subject to a variety of morphosyntactic conditions. Tone-melodic inflectional patterns also play a role in marking tense aspect and clause type, providing the sole distinction between forms such as (Karanga) *akarima* ‘then he plowed’, *ákáríma* ‘he plowed (yesterday)’, and *ákárimá* ‘he having plowed’.

Shona exhibits typical structural properties of Bantu languages. It has a rich system of 19 noun classes marked on nouns by prefixes that encode singular/plural distinctions as well as semantic properties (such as diminutives and augmentatives), with all of the proto-Bantu noun classes represented, except **gũ*- – class 19 **pi*- is represented in the Karanga dialect by the diminutive *svi*-. Noun stems can thus appear in a number of classes, the choice of class marking semantic differences, e.g., *mu-kómaná* ‘boy’, *gómaná* ‘huge boy’, *chi-kómaná* ‘short boy’, *ka-kómaná* ‘tiny sickly boy’, *ru-kómaná* ‘thin boy’, and *svi-kómaná* ‘little boy’. The Bantu class 5 prefix **li*- is itself phonologically null in Shona, and voicing of a stem-initial stop replaces an overt prefix in [ɓángá] ‘knife’ ~ [ma-pángá] ‘knives’, [ɗéndé] ‘gourd’ ~ [ma-téndé] ‘gourds’, [guḍo] ‘baboon’ ~ [ma-kuḍo] ‘baboons’, [jékó] ‘sickle’ ~ [ma-čékó] ‘sickles’: observe that the voiced correspondents of *p*, *t* are the implosives [ɓ, ɗ], not [b, d].

Verb morphology is particularly rich. Stems are composed of a root plus a number of fully and partially productive derivational suffixes marking causative (-*is*-), applicative (-*ir*-), reciprocal (-*an*-), passive (-*w*-), intensive (-*is*-, -*isis*-), reversive (-*Vnur*-), and stative (-*ik*-), as well as reduplication for frequent actions. Pronominal markers include encoding of objects, subjects, and relative clause heads for each of the noun classes. Tense-aspect marking indicates past, present, future, persistent, potential, subjunctive, imperative, hortative, and numerous other distinctions, as well as corresponding negative and subordinate clause forms. The word-formation potential of Shona reaches astronomical levels due to a series of *Aktionsart* prefixes such as -*do*- ‘willingly’, -*ndo*- ‘go’, -*zo*- ‘hypothetical, remote’, -*garo*- ‘always’, -*nyatso*- ‘do well’, and -*raro*- ‘at night’; thus *ndichá-dó-rima* ‘I will gladly plow’, *ndichá-zó-rima* ‘I will plow in the remote future’, and *ndichá-nyatsó-rima* ‘I will

plow well'. These prefixes can be combined and permuted, so that *ndichá-zó-ndó-ráro-rima* 'I will perhaps go plow at night' can also be rendered as *ndichá-ndó-zó-ráro-rima*.

Shona syntax is typical for a Bantu language. Noun phrases are fairly strictly head initial, with noun class agreement governed by the head noun (*vána va-kúru va-viri vá-no* [children big two these] 'these two big children'), though some determiners may be put before the head noun (*iri bángá* [this knife]). Derivational verbal extensions such as the causative and applicative allow a clause to have multiple bare objects, e.g., *ákáríma munda* 'he plowed the field', *ákárímisa Fáraí munda* 'he made Farai plow the field', and *ákárímira munhu munda* 'he plowed the field for the person'. Locative noun phrases can ostensibly function as subjects via inversion (*mu-mbá muw-ákárára vaná* [in-house loc-slept children] the children slept in the house') or passivization

(*mu-mbá muw-ákáfámb-w-á naFáraí* [in-house loc-walk-passive by-Farai] 'in the house was walked by Farai').

Bibliography

- Duramazwi Guru reChiShona*. (2001). Harare: College Press.
- Fivaz D (1970). *Shona morphophonemics and morphology*. Johannesburg: University of Witwatersrand.
- Fortune G (1955). *An analytical grammar of Shona*. London: Longman, Green and Co.
- Fortune G (1980). *Shona grammatical constructions*. Harare: Mercury Press [2nd edn., 1981; 3rd edn., 1985].
- Hannan M (1981). *Standard Shona dictionary* (2nd edn.). Harare: College Press.
- Myers S (1987). Tone and the structure of words in Shona. Doctoral diss., University of Massachusetts, Amherst.
- Odden D (1981). Problems of tone assignment in Shona. Doctoral diss., University of Illinois, Urbana.

Sign Languages

W Sandler, University of Haifa, Haifa, Israel

© 2006 Elsevier Ltd. All rights reserved.

In many ways, sign languages are like spoken languages: They are natural languages that arise spontaneously wherever there is a community of communicators; they effectively fulfill all the social and mental functions of spoken languages; and they are acquired without instruction by children, given normal exposure and interaction. These characteristics have led many linguists to expect sign languages to be similar to spoken languages in significant ways. However, sign languages are different too: As manual-visual languages, sign languages exploit a completely different physical medium from the vocal-auditory system of spoken languages. These two dramatically different physical modalities are also likely to have an effect on the structure of the languages through which they are transmitted.

It is of special interest, then, to compare natural languages in the two modalities. Where the two systems converge, universal linguistic properties are revealed. Where they diverge, the physical medium of transmission is implicated, and its contribution to the form of language in both modalities is illuminated. Neither can be seen quite so clearly if linguists restrict their study to spoken language alone (or to sign language alone). For this and other related reasons, it is often remarked that sign languages provide

us with a natural laboratory for studying the basic characteristics of all human language.

Once the existence of natural language in a second modality is acknowledged, questions such as the following arise: How are such languages born? Are the central linguistic properties of sign languages parallel to those of spoken languages? Is sign language acquired by children in the same stages and time frame as spoken language? Are the same areas of the brain responsible for language in both modalities? What role does modality play in structuring language? In other words, within the architecture of human cognition, do we find the structure of one language 'faculty' or two? Although there is no conclusive answer to this deceptively simple question, an impressive body of research has greatly expanded our understanding of the issues underlying it.

How Do Sign Languages 'Happen'?

Evolution made language possible scores of millennia ago, and there is no human community without it. What sign language teaches us is that humans have a natural propensity for language in two different modalities: vocal-auditory and manual-visual. Since the human ability to use language is so old, and since speech is the predominant medium for its transmission, it seems that spoken languages are either also very old or descended from other languages with a long history. However, sign languages do not have the

same histories as spoken languages because special conditions are required for them to arise and persevere, and for this reason they can offer unique insight into essential features of human language.

The first lesson sign language teaches us is that, given a community of humans, language inevitably emerges. Although we have no direct evidence of the emergence of any spoken language, we can get much closer to the origin of a sign language and, in rare instances, even watch it come into being.

Wherever deaf people have an opportunity to gather and interact regularly, a sign language is born. Typically, deaf people make up a very small percentage of the population (approximately 0.23% in the United States, according to the National Center for Health Statistics, 1994) so that in any given local social group, there may be no deaf people at all or very few of them. The most common setting in which a deaf community can form, then, is a school for deaf children. Such schools only began to be established approximately 200 years ago in Europe and North America. On the basis of this historical information and some reported earlier observations of groups of people using sign language, it is assumed that the oldest extant sign languages do not date back farther than approximately 300 years (Woll *et al.*, 2001). Currently, linguists have the rare opportunity to observe the emergence and development of a sign language from the beginning in a school established in Nicaragua only approximately 25 years ago – an opportunity that is yielding very interesting results.

Graduates of such schools sometimes choose to concentrate in certain urban areas, and wider communities arise and grow, creating their own social networks, institutions, and art forms, such as visual poetry (Padden and Humphries, 2005; Sandler and Lillo-Martin, 2005; Sutton-Spence and Woll, 1999). Deaf society is highly developed in some places, and the term ‘Deaf’ with a capital D has come to refer to members of a minority community with its own language and culture rather than to people with an auditory deficit.

It is not only the genesis of a sign language that is special; the way in which it is passed down from generation to generation is unusual as well. Typically, fewer than 10% of deaf children acquire sign language from deaf parents, and of those deaf parents, only a small percentage are native signers. The other 90+% of deaf children have hearing parents and may only be exposed to a full sign language model when they go to school. These social conditions taken together with certain structural properties of sign languages have prompted some linguists to compare them to spoken creoles (Fischer, 1978).

Another way in which a deaf social group and concomitant sign language can form is through the propagation of a genetic trait within a small village or town through consanguineous marriage, resulting in a proportionately high incidence of deafness and the spread of the sign language among both deaf and hearing people. Potentially, this kind of situation can allow us to observe the genesis and development of a language in a natural community setting. Although the existence of such communities has been reported occasionally (see Groce, 1985), no comprehensive linguistic description of a language arising in such a community has been provided.

These, then, are the ways in which sign languages happen. The existence of many sign languages throughout the world – the number 103 found in the Ethnologue database is probably an underestimate – confirms the claim that the emergence of a highly structured communication system among humans is inevitable. If the oral–aural channel is unavailable, language springs forth in the manual–visual modality.

Not only does such a system emerge in the absence of audition, but its kernels can be also observed even in the absence of both a community and a language model. Deaf children who live in hearing households in which only oral language is used, who have not yet experienced speech training, and thus have no accessible language model, devise their own systematic means of communication called home sign, studied in exquisite detail by Goldin-Meadow and colleagues (Goldin-Meadow, 2003). The gesture talk of these children contains the unmistakable imprint of a real linguistic system, and as such it offers a unique display of the fundamental human genius for language.

At the same time, the form and content of home sign are rudimentary and do not approach the richness and complexity of a language used by a community, spoken or signed. This confronts us with another important piece of information: Language as we know it is a social phenomenon. Although each brain possesses the potential for language, it takes more than one brain to create a complex linguistic system.

The Linguistic Structure of Sign Language

Hearing people use gesture, pantomime, and facial expression to augment spoken language. Naturally, the ingredients of these forms of expression are available to sign languages too. The apparent familiarity of the raw material that contributes to the formation of sign languages has led many a naïve observer to the mistaken assumption that sign languages are actually simple gesture systems. However, instead of forming an idiosyncratic, ancillary system like the one that accompanies speech, these basic ingredients contribute

to a primary linguistic system in the creation of a sign language, a system with many of the same properties found in spoken languages. In fact, linguistic research has demonstrated that there are universal organizing principles that transcend the physical modality, subsuming spoken and signed languages alike.

The Phonology of Sign Language

William Stokoe (1960) demonstrated that the signs of American Sign Language (ASL) are not gestures: They are not holistic icons. Instead, Stokoe showed that they are composed of a finite list of contrastive meaningless units like the phonemes of spoken languages. These units combine in constrained ways to create the words of the language. Although there are some differences among different sign languages in their phonological inventories and constraints, there are many common properties, and the generalizations presented here hold across sign languages, unless otherwise indicated.

Stokoe established three major phonological categories: hand shape, location, and movement. Each specification within each of the three major categories was treated as a phoneme in Stokoe's work. Later researchers accepted these categories but proposed that the specifications within each category function not as phonemes but as phonological features. The ASL signs SICK and TOUCH, illustrated in Figure 1,

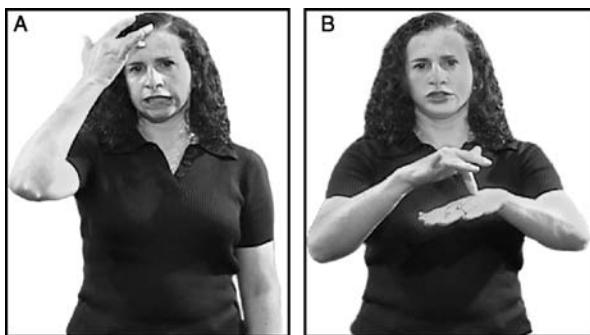


Figure 1 ASL minimal pair distinguished by a location feature. (A) SICK and (B) TOUCH.

have the same hand shape and the same straight movement. They are distinguished by location only: The location for SICK is the head, whereas the location for TOUCH is the nondominant hand. Minimal pairs such as this one, created by differences in one feature only, exist for the features of hand shape and movement as well. Although the origins of these and other (but certainly not all) signs may have been holistic gestures, they have evolved into words in which each formational element is contrastive but meaningless in itself.

Two other defining properties of phonological systems exist in sign languages as well: constraints on the combination of phonological elements and rules that systematically alter their form. One phonological constraint on the form of a (monomorphemic) sign concerns the set of two-handed signs. If both hands are involved, and if both hands also move in producing the sign (unlike TOUCH, in which only one hand moves), then the two hands must have the same hand shape and the same (or mirror) location and movement (Battison, 1978). An example is DROP, shown in Figure 2B: Both hands move, and they are identical in all other respects as well. The second defining property, changes in the underlying phonological form of a sign, is exemplified by hand shape assimilation in compounds. In one lexicalized version of the ASL compound, MIND + DROP = FAINT, the hand shape of the first member, MIND, undergoes total assimilation to the hand shape of the second member, DROP, as shown in Figure 2.

Stokoe believed that hand shapes, locations, and movements cooccur simultaneously in signs, an internal organization that differs from the sequentiality of consonants and vowels in spoken language. Liddell (1984) took exception to that view, showing that there is phonologically significant sequentiality in this structure. Sandler (1989) further refined that position, arguing that the locations (L) and movement (M) within a sign are sequentially ordered, whereas the hand configuration (HC) is autosegmentally associated to these elements – typically, one hand



Figure 2 Hand configuration assimilation in the ASL compound. (A) MIND + (B) DROP = (C) FAINT.

configuration (i.e., one hand shape with its orientation) to a sign, as shown in the representation in **Figure 3**. The first location of the sign TOUCH in **Figure 1B**, for example, is a short distance above the nondominant hand, the movement is a straight path, and the second location is in contact with the nondominant hand. The hand shape of the whole sign is ✋ .

Under assimilation, as in **Figure 2**, the HC of the second member of the compound spreads regressively to the first member in a way that is temporally autonomous with respect to the Ls and Ms, manifesting the autosegmental property of stability (Goldsmith, 1979). The sequential structure of signs is still a good deal more limited than that of words in most spoken languages, however, usually conforming to this canonical LML form even when the signs are morphologically complex (Sandler, 1993).

Morphology

All established sign languages studied to date, like the overwhelming majority of spoken languages, have complex morphology. First, as shown in **Figure 2**, compounding is very common. In addition, some sign languages have a limited number of sequential affixes. For example, Israeli Sign Language (ISL) has a derivational negative suffix, similar in meaning to English *-less*, that was grammaticalized from a free word glossed NOT-EXIST. This suffix has two allomorphs, depending on the phonological form of the base, illustrated in **Figure 4**. If the base is two-handed, so is the suffix, whereas one-handed bases trigger the one-handed allomorph of the suffix.

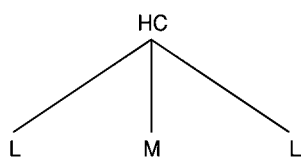


Figure 3 The canonical form of a sign. From Sandler (1989).

Sign languages typically have a good deal of complex morphology, but most of it is not sequential like the examples in **Figures 3** and **4**. Instead, signs gain morphological complexity by simultaneously incorporating morphological elements (Fischer and Gough, 1978). The prototypical example, first described in detail in ASL (Padden, 1988) but apparently found in all established sign languages, is verb agreement. This inflectional system is prototypical not only because of the simultaneity of structure involved but also because of its use of space as a grammatical organizing property.

The system relies on the establishment of referential loci – points on the body or in space that refer to referents in a discourse – that might be thought of as the scaffolding of the system. In **Figure 5**, loci for first person and third person are established.

In the class of verbs that undergoes agreement, the agreement markers correspond to referential loci established in the discourse. Through movement of the hand from one locus to the other, the subject is marked on the first location of the verb and the object on the second. **Figure 6A** shows agreement for the ASL agreeing verb, ASK, where the subject is first person and the object is third person. **Figure 6B** shows the opposite: third person subject and first person object. The requirement that verb agreement must refer independently to the first and last locations in a sign was one of the motivations for Liddell's (1984) claim that signs have sequential structure.

Although each verb in **Figure 6** includes three morphemes, each still conforms to the canonical LML form shown in **Figure 3**. The agreement markers are encoded without sequential affixation. Sign language verb agreement is a linguistic system, crucially entailing such grammatical concepts as coreference, subject and object, and singular and plural. It is also characterized by sign language-specific properties, such as the restriction of agreement to a particular class of verbs (Padden, 1988), identified mainly on semantic grounds (Meir, 2002).

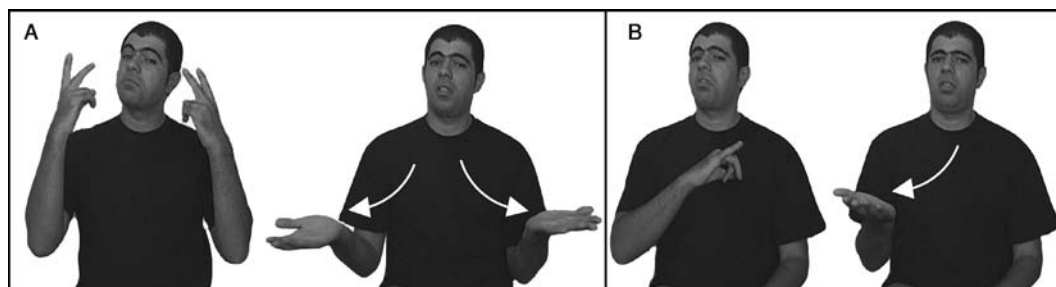


Figure 4 Allomorphs of an ISL suffix. (A) IMPORTANT-NOT-EXIST (without importance). (B) INTERESTING-NOT-EXIST (without interest).

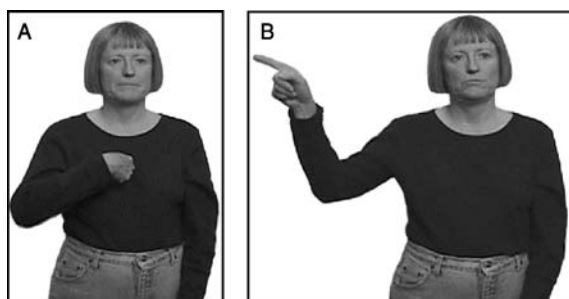


Figure 5 Referential loci. (A) First person. (B) Third person.

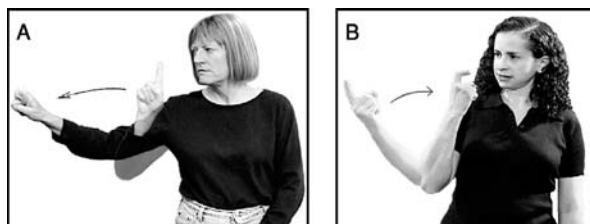


Figure 6 Verb agreement. (A) 'I ask him/her.' (B) 's/he asks me.'

Another productive inflectional morphological system found across sign languages is temporal and other aspectual marking, in which the duration of Ls and Ms, the shape of the movement path, or both may be altered, and the whole form may be reduplicated, to produce a range of aspects, such as durational, continuative, and iterative (Klima and Bellugi, 1979). This system has templatic characteristics, lending itself to an analysis that assumes CV-like LM templates and nonlinear associations of the kind McCarthy (1981) proposed for Semitic languages (Sandler, 1989, 1990).

Figure 4 demonstrated that some limited sequential affixation exists in sign languages. However, the most common form of sign language words by far, whether simple or complex, is LML (setting aside gemination of Ls and Ms in the aspectual system, which adds duration but not segmental content). In fact, even lexicalized compounds such as the one shown in Figure 2 often reduce to this LML form. If movement (M) corresponds to a syllable nucleus in sign language, as Perlmutter (1992), Brentari (1998), and others have argued, then it appears that monosyllabicity is ubiquitous in ASL (Coulter, 1982) and in other sign languages as well. In the midst of a morphological system with many familiar linguistic characteristics (e.g., compounding, derivational morphology, inflectional morphology, and allomorphy), we see in the specific preferred monosyllabic form of sign language words a clear modality effect (Sandler and Lillo-Martin, 2005).

No overview of sign language morphology would be complete without a description of the classifier



Figure 7 Classifier construction in ASL.

subsystem. This subsystem is quintessentially 'sign language,' exploiting the expressive potential of two hands forming shapes and moving in space, and molding it into a linguistic system (Emmorey, 2003; Supalla, 1986). Sign languages use classifier constructions to combine physical properties of referents with the spatial relations among them and the shape and manner of movement they execute. In this subsystem, there is a set of hand shapes that classify referents in terms of their size and shape, semantic properties, or other characteristics in a classificatory system that is reminiscent of verbal classifiers found in a variety of spoken languages (Senft, 2002). These hand shapes are the classifiers that give the system its name. An example of a classifier construction is shown in Figure 7. It describes a situation in which a person is moving ahead, pulling a recalcitrant dog zigzagging behind. The upright hand shape embodies an upright human classifier and the legged hand shape a legged creature.

What is unusual about this subsystem is that each formational element – the hand shape, the location, and the movement – has meaning. That is, each has morphological status. This makes the morphemes of classifier constructions somewhat anomalous since sign language lexicons are otherwise built of morphemes and words in which each of these elements is meaningless and has purely phonological status. Furthermore, constraints on the cooccurrence of these elements in other lexical forms do not hold on classifier constructions. In Figure 7, for example, the constraint on interaction of the two hands described in the section on phonology is violated. Zwitserlood (2003) suggested that each hand in such classifier constructions is articulating a separate verbal constituent, and that the two occur simultaneously – a natural kind of structure in sign language and found universally in them, but one that is inconceivable in spoken language. Once again, sign language presents a conventionalized system with linguistic properties, some familiar from spoken languages and some modality driven.

Syntax

As in other domains of linguistic investigation, the syntax of sign languages displays a large number of characteristics found universally in spoken languages. A key example is recursion – the potential to repeatedly apply the same rule to create sentences of ever increasing complexity – argued to be the quintessential linguistic property setting human language apart from all other animal communication systems (Hausser *et al.*, 2002). Specifically, through embedding or conjoining, recursion can result in sentences of potentially infinite length. These two different ways of creating complex sentences have been described and distinguished from one another in ASL. For example, a process that tags a pronoun that is coreferential with the subject of a clause onto the end of a sentence may copy the first subject in a string, only if the string contains an embedded clause, but not if the second clause is coordinate (Padden, 1988). In example (1), the subscripts stand for person indices marked through verb agreement, and INDEX is a pointing pronominal form, here a pronoun copy of the matrix subject, MOTHER. (These grammatical devices were illustrated in Figures 5 and 6.)

- (1a) MOTHER SINCE_i PERSUADE_j SISTER_i
 COME_i INDEX_j
 ‘My mother has been urging my sister to come
 and stay here, she (mother) has.’
- (1b) *_iHIT_j INDEX_i TATTLE MOTHER_i INDEX_j.
 ‘I hit him and he told his mother, I did.’

The existence of strict constraints on the relations among nonadjacent elements and their interpretation is a defining characteristic of syntax. A different category of constraints of this general type concerns movement of constituents from their base-generated position, such as the island constraints first put forward by Ross (1967) and later subsumed by more general constraints. One of these is the WH island constraint, prohibiting the movement of an element out of a clause with an embedded WH question. The sentence, *Lynn wonders [what Jan thinks]* is okay, but the sentence **It’s Jan that Lynn wonders [what — thinks]* is ruled out. Lillo-Martin (1991) demonstrated that ASL obeys the WH island constraint with the sentences shown in example (2). Given the relative freedom of word order often observed in sign languages such as ASL, it is significant that this variability is nevertheless restricted by universal syntactic constraints.

- (2a) PRO DON’T-KNOW [‘WHAT’ MOTHER
 LIKE].
 ‘I don’t know what Mom likes.’
 _____t

- (2b) MOTHER, PRO DON’T KNOW [‘WHAT’
 ______i LIKE].
 * ‘As for Mom, I don’t know what likes.’

The line over the word MOTHER in (2b) indicates a marker that is not formed with the hands, in this case a backward tilt of the head together with raised eyebrows, marking the constituent as a topic (t) in ASL. There are many such markers in sign languages, which draw from the universal pool of idiosyncratic facial expressions and body postures available to all human communicators. These expressions and postures become organized into a grammatical system in sign languages.

A Grammar of the Face

When language is not restricted to manipulations of the vocal tract and to auditory perception, it is free to recruit any parts of the body capable of rapid, variegated articulations that can be readily perceived and processed visually, and so it does. All established sign languages that have been investigated use non-manual signals – facial expressions and head and body postures – grammatically. These expressions are fully conventionalized and their distribution is systematic.

Early research on ASL showed that certain facial articulations, typically of the mouth and lower face, function as adjectivals and as manner adverbials, the latter expressing such meanings as ‘with relaxation and enjoyment’ and ‘carelessly’ (Liddell, 1980). Other sign languages have been reported to use lower face articulations in similar ways. The specific facial expressions and their associated meanings vary from sign language to sign language. Figure 8 gives examples of facial expressions of this type in ASL, ISL, and British Sign Language.

A different class of facial articulations, particularly of the upper face and head, predictably cooccur with specific constructions, such as yes/no questions, WH questions, and relative clauses in ASL and in many other sign languages. Examples from ISL shown in Figure 9A illustrate a yes/no question (raised brows, wide eyes, and head forward), Figure 9B a WH question (furrowed brows and head forward), and Figure 9C the facial expression systematically

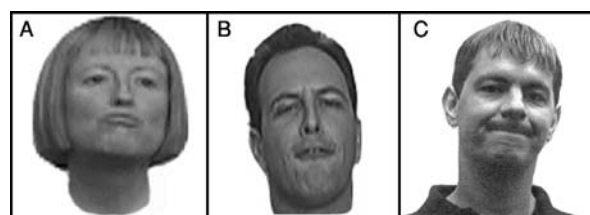


Figure 8 Lower face articulations. (A) ASL ‘with relaxation and enjoyment.’ (B) ISL ‘carefully.’ (C) BSL ‘exact.’



Figure 9 Upper face articulations. (A) yes/no question, (B) WH question, and (C) 'shared information.'

associated in that language with information designated as 'shared' for the purpose of the discourse (squinted eyes). Although some of these facial articulations may be common across sign languages (especially those accompanying yes/no and WH questions), these expressions are not iconic. Some researchers have proposed that they evolved from more general affective facial expressions associated with emotions. In sign languages, however, they are grammaticalized and formally distinguishable from the affective kind that signers, of course, also use (Reilly *et al.*, 1990).

Observing that nonmanual signals of the latter category often cooccur with specific syntactic constructions, Liddell (1980) attributed to them an expressly syntactic status in the grammar of ASL, a view that other researchers have adopted and expanded (Neidle *et al.*, 2000; Petronio and Lillo-Martin, 1997). A competing view proposes that they correspond to intonational tunes (Reilly *et al.*, 1990) and participate in a prosodic system (Nespor and Sandler, 1999). Wilbur (2000) presented evidence that nonmanuals convey many different kinds of information – prosodic, syntactic, and semantic. A detailed discussion can be found in Sandler and Lillo-Martin (2005).

Acquisition of Sign Language

Nowhere is the 'natural laboratory' metaphor more appropriate than in the field of sign language acquisition. This area of inquiry offers a novel and especially revealing vantage point from which to address weighty questions about the human capacity for language. Research has shown, for example, that children acquire sign language without instruction, just as hearing children acquire spoken language, and according to the same timetable (Newport and Meier, 1985). These findings lend more credence to the view, established by linguistic research on the adult system, that languages in the two modalities share a significant amount of cognitive territory; children come equipped for the task of acquiring language in either modality equally.

Studies have also shown that signing children attend to grammatical properties, decomposing and

overgeneralizing them as they advance through the system, sometimes even at the expense of the iconic properties inherent in that system. For example, even the pointing gesture used for pronouns (see Figure 5) is analyzed as an arbitrary grammatical element by small children, who may go through a stage in which they make mistakes, pointing at 'you' to mean 'me' (Pettito, 1987). Meier (1991) discovered counter-iconic errors in verb agreement (see Figure 6), similarly indicating that children are performing a linguistic analysis, exploring spatial loci as grammatical elements and not as gestural analogues to actual behavior and events.

Due to the social conditions surrounding its acquisition, sign language lends novel insight into two key theories about language and its acquisition: the critical period hypothesis and the notion that the child makes an important contribution to the crystallization of a grammar. Some deaf children are raised in oral environments, gaining access to sign language later in life. Studies comparing the ASL performance of early and late learners found that the age of exposure is critical for acquisition of the full grammatical system (Newport, 1990) and its processing (Mayberry and Eichen, 1991), providing convincing support for Lenneberg's (1967) critical period hypothesis. An untainted perspective on the child's contribution can be gained where the input to the child is simpler and/or less systematic than a full language system, as with pidgins (Bickerton, 1981). Researchers of the sign language that originated in the Nicaraguan school mentioned previously studied the communication system conventionalized from the home sign brought to the school by the first cohort of children. This system served as input to the second cohort of children younger than the age of 10 years who later arrived at the school. Comparing the language of the two cohorts, the researchers found that children make an important contribution: The second cohort of signers developed a significantly more structured and regular system than the one that served as their input (Kegl *et al.*, 1999; Senghas *et al.*, 2004).

Sign Language and the Brain

Broadly speaking, it is established that most spoken language functions involve extensive activity in specific areas of the left hemisphere of the brain, whereas much of visuospatial cognition involves areas of the right cerebral hemisphere. Therefore, the discovery that sign language, like spoken language, is primarily controlled by the left hemisphere despite its exploitation of the visuospatial domain is striking and significant (Emmorey, 2002; Poizner *et al.*, 1987). Various

explanations for left hemisphere dominance for language are currently on the table, such as the more general ability of the left hemisphere to process rapidly changing temporal events (Fitch *et al.*, 1997). This explanation has been rejected for sign language by some researchers on the grounds that sign language production is slower than that of spoken language (Hickock *et al.*, 1996). Whatever explanation is ultimately accepted, Emmorey (2002) and others have argued that similarities in the kind of cognitive operations inherent in the organization and use of language in the two modalities should not be ignored in the search.

Although most language functions are controlled by the left hemisphere, some do show right hemisphere involvement or advantage. With respect to sign language, there is evidence that the right hemisphere may be more involved in producing and comprehending certain topographic/spatial aspects of sign language, particularly those involving classifier constructions (Emmorey *et al.*, 2002). Although this result sits well with the known right hemisphere advantage for spatial processing, it is made even more interesting when added to discoveries of right hemisphere dominance for certain other spoken and sign language functions that may be related to the classifier system, such as processing words with imageable, concrete referents (Day, 1979; Emmorey and Corina, 1993). Findings such as these are an indication of the way in which sign language research adds important pieces to the puzzle of language organization in the brain.

Language Modality, Language Age, and the Dinner Conversation Paradox

A large body of research, briefly summarized here, attributes to sign languages, essential linguistic properties that are found in spoken languages as well (Sandler and Lillo-Martin, 2005). Also, like different spoken languages, sign languages are not mutually intelligible. A signer of ISL observing a conversation between two signers of ASL will not understand it. Although cross sign language research is in its infancy, some specific linguistic differences from sign language to sign language have already been described.

At the same time, there is a rather large group of predictable similarities across sign languages and, as Newport and Supalla (2000: 109) stated, “A long dinner among Deaf users of different sign languages will, after a while, permit surprisingly complex interchanges.” Here we find a difference between signed and spoken languages: One would hardly expect even the longest of dinners to result in complex interchanges among monolingual speakers of English

and Mandarin Chinese. Although it is clear that more differences across sign languages will be uncovered with more investigation and more sophisticated research paradigms, it is equally certain that the dinner conversation paradox will persist. Two reasons have been suggested for crosssign language similarities: the effect of modality on language structure and the youth of sign languages.

Modality Effects

Modality is responsible for two interwoven aspects of sign language form, both of which may contribute to similarities across sign languages: (i) an iconic relation between form and meaning, and (ii) simultaneity of structure. Because the hands can represent physical properties of concrete objects and events iconically, this capability is abundantly exploited in sign languages, both in lexical items and in grammatical form. Although spoken languages exhibit some iconicity in onomatopoeia, ideophones, and the like, the vocal-auditory medium does not lend itself to direct correspondence between form and meaning so that the correspondence in spoken language is necessarily more arbitrary.

Iconicity in Sign Language Leafing through a sign language dictionary, one immediately notices the apparent iconicity of many signs. An example is the ISL sign for BOOK, shown in Figure 10, which has the appearance of a book opening. Although clearly widespread, iconicity in sign language must be understood in the right perspective. Many signs are not iconic or not obviously motivated, among them the signs for abstract concepts that exist in all sign languages. Interestingly, even the iconicity of signs that are motivated is not nearly so apparent to nonsigners if the translations are not available (Klima and Bellugi, 1979). In addition, the presence of iconicity in sign language does not mean that their vocabularies are overwhelmingly similar to one another. In fact, although even unrelated sign languages have some overlap in vocabulary due to motivatedness,



Figure 10 An iconic sign: (ISL) BOOK.

their vocabularies are much more different from one another than one might expect (Currie *et al.*, 2002). Nevertheless, the kind of symbolization and metaphorical extension involved in creating motivated signs may be universal (Taub, 2001). For example, a bird is represented in ISL with a sign that looks like wings and in ASL with a sign that looks like a beak, and experience with this kind of symbolization in either sign language may make such signs easier to interpret in the other.

Simultaneity in Sign Languages Another modality feature is simultaneity of structure, alluded to previously. Some researchers have argued that constraints on production, perception, and short-term memory conspire to create simultaneity of linguistic structure (Bellugi and Fischer, 1972; Emmorey, 2002). Interestingly, iconicity also makes a contribution to simultaneity of structure, especially when one looks beyond the lexicon to grammatical forms of a more complex nature.

The hands moving in space are capable of representing events that simultaneously involve a predicate and its arguments (e.g., giving something to someone or skimming across a bumpy surface in a car) with a form that is correspondingly simultaneous. The result is verb agreement (exemplified in **Figure 6**) and classifier constructions (exemplified in **Figure 7**). Therefore, these structures, with particular grammatical properties, are found in all established sign languages that have been studied, leading to the observation that sign languages belong to a single morphological type (Aronoff *et al.*, 2005). Although the grammatical details of this morphology differ from sign language to sign language, the principles on which they are based are the same, and this similarity makes another welcome contribution at the dinner table.

The Role of Language Age

Recent work pinpoints the role of language age in the structure of sign language, indicating how age may be partly responsible for the impression that crosssign language differences are less abundant than is the case across spoken languages. It does so by comparing the type of morphology ubiquitously present in sign languages with a language-specific type (Aronoff *et al.*, 2005). This study noted that the form taken by the verb agreement and classifier systems in all established sign languages is similar (although not identical) due to the modality pressures of iconicity and simultaneity sketched previously, but that sequential affixes of the kind exemplified in **Figure 4** vary widely between the sign languages studied. Such affixes, arbitrary rather than iconic in form and limited in

number, develop through grammaticalization processes, and these processes take time. Given time, more such arbitrary, sign language-specific processes are predicted to develop.

The physical channel of transmission affects language in both modalities. Where sign languages are more simultaneously structured, spoken languages are more linear. Where spoken languages are mostly arbitrary, sign languages have a good deal of iconicity. However, none of these qualities are exclusive to one modality; it is only a matter of degree.

Bibliography

- Aronoff M, Meir I & Sandler W (2005). 'The paradox of sign language morphology.' *Language*. 301–344.
- Battison R (1978). *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstok.
- Bellugi U & Fischer S (1972). 'A comparison of sign language and spoken language.' *Cognition* 1, 173–200.
- Bellugi U & Klima E S (1982). 'The acquisition of three morphological systems in American Sign Language.' *Papers and Reports on Child Language Development* 21, 1–35.
- Bickerton D (1981). *The roots of language*. Ann Arbor, MI: Karoma.
- Brentari D (1998). *A prosodic model of sign language phonology*. Cambridge: MIT Press.
- Coulter G (1982). 'On the nature of ASL as a monosyllabic language.' Paper presented at the annual meeting of the Linguistic Society of America, San Diego.
- Currie A M, Meier R & Walters K (2002). 'A crosslinguistic examination of the lexicons of four signed languages.' In Meier R, Cormier K & Quinto-Pozos D (eds.) *Modality and structure in signed language and spoken language*. Cambridge, UK: Cambridge University Press. 224–236.
- Day J (1979). 'Visual half-field word recognition as a function of syntactic class and imageability.' *Neuropsychologia* 17, 515–520.
- Emmorey K (2002). *Language, cognition, and the brain: insights from sign language research*. Mahwah, NJ: Erlbaum.
- Emmorey K (ed.) (2003). *Perspectives on classifier constructions in sign languages*. Mahwah, NJ: Erlbaum.
- Emmorey K & Corina D P (1993). 'Hemispheric specialization for ASL signs and English words: differences between imageable and abstract forms.' *Neuropsychologia* 31, 645–653.
- Emmorey K & Lane H (eds.) (2000). *The signs of language revisited: an anthology to honor Ursula Bellugi and Edward Klima*. Mahwah, NJ: Erlbaum.
- Emmorey K, Damasio H, McCullough S, Grabowski T, Ponto L, Hichwa R & Bellugi U (2002). 'Neural systems underlying spatial language in American Sign Language.' *NeuroImage* 17, 812–824.
- Fischer S (1978). 'Sign language and creoles.' In Siple P (ed.) *Understanding language through sign language research*. New York: Academic Press. 309–331.

- Fischer S & Gough B (1978). 'Verbs in American Sign Language.' *Sign Language Studies* 7, 17–48.
- Fitch R H, Miller S & Tallal P (1997). 'Neurobiology of speech perception.' *Annual Review of Neuroscience* 20, 331–353.
- Goldin-Meadow S (2003). *The resilience of language: what gesture creation in deaf children can tell us about how all children learn language*. New York: Psychology Press.
- Goldsmith J (1979). *Autosegmental phonology*. New York: Garland.
- Groce N E (1985). *Everyone here spoke sign language: hereditary deafness on Martha's vineyard*. Harvard, MA: Harvard University Press.
- Hausser M D, Chomsky N & Fitch W T (2002). 'The faculty of language: what is it, who has it, and how did it evolve?' *Science* 29(5598), 1569–1579.
- Hickock G, Klima E S & Bellugi U (1996). 'The neurobiology of signed language and its implications for the neural basis of language.' *Nature* 381, 699–702.
- Kegl J, Senghas A & Coppola M (1999). 'Creation through contact: sign language emergence and sign language change in Nicaragua.' In DeGraff M (ed.) *Language creation and language change: creolization, diachrony, and development*. Cambridge: MIT Press. 179–237.
- Klima E S & Bellugi U (1979). *The signs of language*. Cambridge, MA: Harvard University Press.
- Kyle J & Woll B (1985). *Sign language. The study of deaf people and their language*. Cambridge, UK: Cambridge University Press.
- Lenneberg E (1967). *Biological foundations of language*. New York: Wiley.
- Liddell S (1980). *American Sign Language syntax*. The Hague, The Netherlands: Mouton.
- Liddell S (1984). THINK and BELIEVE: sequentiality in American Sign Language. *Language* 60, 372–399.
- Lillo-Martin D (1991). *Universal grammar and American Sign Language: setting the null argument parameters*. Dordrecht, The Netherlands: Kluwer Academic.
- Mayberry R & Eichen E (1991). 'The long-lasting advantage of learning sign language in childhood: another look at the critical period for language acquisition.' *Journal of Memory and Language* 30, 486–512.
- McCarthy J (1981). 'A prosodic theory of nonconcatenative morphology.' *Linguistic Inquiry* 20, 71–99.
- Meier R (1991). 'Language acquisition by deaf children.' *American Scientist* 79, 60–70.
- Meir I (2002). 'A cross-modality perspective on verb agreement.' *Natural Language and Linguistic Theory* 20(2), 413–450.
- Neidle C, Kegl J, MacLaughlin D, Bahan B & Lee R G (2000). *The syntax of American Sign Language: functional categories and hierarchical structure*. Cambridge: MIT Press.
- Nespor M & Sandler W (1999). 'Prosody in Israeli Sign Language.' *Language and Speech* 42(2–3), 143–176.
- Newport E L (1990). 'Maturation constraints on language learning.' *Cognitive Science* 14, 11–28.
- Newport E L & Meier R P (1985). 'The acquisition of American Sign Language. In Slobin D (ed.) *The cross-linguistic study of language acquisition*. Hillsdale, NJ: Erlbaum. 881–938.
- Newport E L & Supalla T (2000). 'Sign language research at the millennium.' In Emmorey & Lane (eds.), 103–114.
- Padden C (1988). *Interaction of morphology and syntax in American Sign Language*. New York: Garland.
- Padden C & Humphries T (2005). *Inside deaf culture*. Cambridge, MA: Harvard University Press.
- Perlmutter D (1992). 'Sonority and syllable structure in American Sign Language. *Linguistic Inquiry* 23(3), 407–442.
- Pettito L A (1987). 'On the autonomy of language and gesture: evidence from the acquisition of personal pronouns in American Sign Language. *Cognition* 27, 1–52.
- Petronio K & Lillo-Martin D (1997). 'WH-movement and the position of spec-CP: evidence from American sign Language.' *Language* 73(1), 18–57.
- Poizner H, Klima E S & Bellugi U (1987). *What the hands reveal about the brain*. Cambridge: MIT Press.
- Reilly J S, McIntire M & Bellugi U (1990). 'The acquisition of conditionals in American Sign Language: grammaticized facial expressions.' *Applied Psycholinguistics* 11(4), 369–392.
- Ross J (1967). Constraints on variables in syntax. Unpublished doctoral diss., MIT.
- Sandler W (1989). *Phonological representation of the sign: linearity and nonlinearity in American Sign Language*. Dordrecht, The Netherlands: Foris.
- Sandler W (1990). 'Temporal aspects and ASL phonology.' In Fischer S D & Siple P (eds.) *Theoretical issues in sign language research*. Chicago: University of Chicago Press. 7–35.
- Sandler W (1993). 'A sonority cycle in American Sign Language.' *Phonology* 10(2), 243–279.
- Sandler W & Lillo-Martin D (2005). *Sign language and linguistic universals*. Cambridge, UK: Cambridge University Press.
- Senft G (2002). *Systems of nominal classification*. Cambridge, UK: Cambridge University Press.
- Senghas A, Kita S & Ozyurek A (2004). 'Children creating core properties of language: evidence from an emergence sign language in Nicaragua.' *Science* 305, 1779–1782.
- Stokoe W (1960). *Sign language structure: an outline of the visual communication systems of the American Deaf*. Studies in linguistics, occasional papers No. 8. Silver Spring, MD: Linstok.
- Supalla T (1986). 'The classifier system in American Sign Language.' In Craig C (ed.) *Noun classification and categorization*. Philadelphia: Benjamins. 181–214.
- Sutton-Spence R & Woll B (1999). *The linguistics of British Sign Language: an introduction*. Cambridge, UK: Cambridge University Press.
- Taub S (2001). *Language from the body: iconicity and metaphor in American Sign Language*. Cambridge, UK: Cambridge University Press.
- Wilbur R B (2000). 'Phonological and prosodic layering of non-manuals in American Sign Language.' Emmorey & Lane (eds.), 215–243.

Woll B, Sutton-Spence R & Elton F (2001). 'Multilingualism: the global approach to sign language.' In Lucas C (ed.) *The sociolinguistics of sign languages*. Cambridge, UK: Cambridge University Press. 8–32.

Zwitserlood I (2003). Classifying hand configuration in Nederlandse Gebarentaal Doctoral diss., University of Utrecht, Holland Institute of Linguistics.

Sign Language: Morphology

T Johnston, Macquarie University, Sydney, Australia

© 2006 Elsevier Ltd. All rights reserved.

Morphology deals with the regular, minimal, meaning-bearing units in language – morphemes – which are words or parts of words. Morphemes can effect changes in meaning by signaling the creation of a new word or a change in word class (derivation), or by signaling grammatical information such as case, number, person, aspect, tense, etc., (inflection).

Individual signs in a signed language are the basic equivalent of words in a spoken language. Each signed language has a vocabulary of conventional lexical signs which are often monomorphemic. In the closely related Australian and British Sign Languages (Auslan and BSL, respectively), for example, none of the formational aspects of the sign **SISTER** has any separate meaning of its own (see **Sign Languages of the World**). See **Figure 1**.

The type of morphological processes commonly found in signed languages seems to be influenced by the fact that most lexical signs are monosyllabic or, at most, bisyllabic (Johnson and Liddell, 1986; Liddell, 1984; Sandler, 1995; Wilbur, 1993). Signed languages appear to favor simultaneous sign internal modification, rather than the concatenation of morphemes. This may be related to the fact that the larger size of the articulators in signed languages (the hands, arms, face, and body) means that each sign gesture takes more time to execute than each

spoken articulatory gesture (Bellugi and Fischer, 1972). If segments are added to a stem, producing a multisyllabic sign, processes of assimilation and deletion tend to restructure the resulting sign into a bisyllabic or monosyllabic one with simultaneously expressed morphemes.

This process is most clearly seen in the formation of new signs through compounding. In Auslan/BSL, for example, the sign **CHECK** derives from **SEE** and **MAYBE**. **SEE** has lost its outward movement with final hold and has incorporated the anticipated handshake of **MAYBE**, while **MAYBE** has lost its repeated twisting movement. The compound has a single syllable (Sutton-Spence and Woll, 1999). See **Figure 2**.

With the exception of prefixes in Israeli Sign Language (Aronoff *et al.*, 2000), the few affixes that have been identified in signed languages are all suffixes. Indeed, many appear to have grammaticized from one of the elements in erstwhile compounds. For example, a negative suffix –NEG can be attached to **AGREE** to derive the new sign **DISAGREE**, in Auslan/BSL. The affix appears to be a reduced form of an independent sign, which itself seems related to a gesture (meaning something like 'not know') shared with the hearing culture. A similar suffix (in both form and meaning) has been identified in a number of signed languages (Zeshan, 2006). See **Figure 3**.

Researchers have also identified a derivational process whereby stem signs for certain units that are enumerable (e.g., **TOMORROW**, **WEEK**) may incorporate numeral handshapes to create specific signs for



SISTER

Figure 1 A monomorphemic sign. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

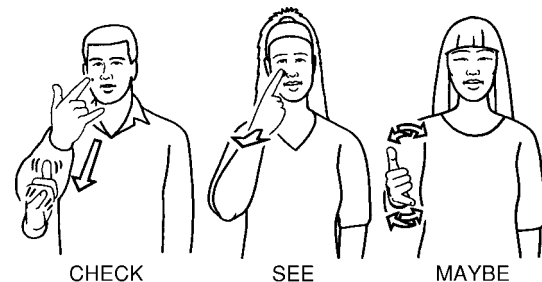


Figure 2 A compound and its components. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

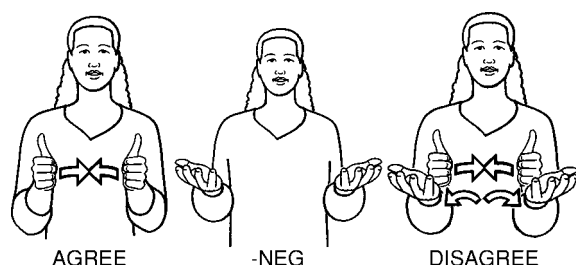


Figure 3 Derivation through negative affixation. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

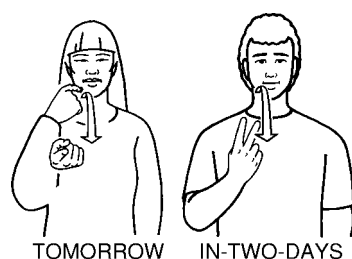


Figure 4 Number incorporation in Auslan/BSL. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

specific numbers of these units (e.g., IN TWO DAYS, TWO-WEEKS). See **Figure 4**.

Modification of the quality of the movement in a sign can also be used to derive new signs, such as BUSY from WORK (Auslan/BSL). A similar process derives NARROW-MINDED/PRUDISH from CHURCH in ASL, and this has been compared to templated morphology as found in Semitic languages (Fernald and Napoli, 2000). See **Figure 5**.

It is often difficult to clearly distinguish between stem modification or suprasegmental modification in signed languages. In many respects, modifying the movement parameter of a sign is akin to changing a vowel (a stem modification); however, in others, modifying for manner of movement is akin to tone (a suprasegmental modification). The derivation of nouns from verbs in some signed languages is a case in point.

Originally described in ASL (Supalla and Newport, 1978), this morphological process, in which nouns are derived from verbs by a sign-internal modification of movement, also has parallels in other signed languages. The continuous single movement found in the verb is modified to be restrained and tense, and often repeated, in the noun, as in the Auslan/BSL pair DOOR and OPEN-DOOR. See **Figure 6**.

However, typical exemplars in many signed languages overlap considerably in both form and

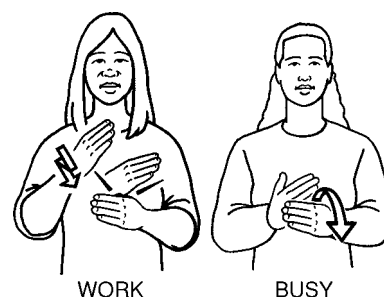


Figure 5 Derivation through movement modification. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

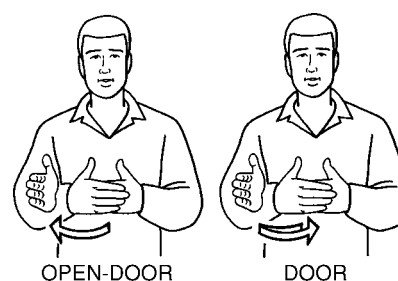


Figure 6 The derivation of a nominal from a verb. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

meaning, and the productivity of the derivational process appears influenced by underlying iconicity. Though it is to be expected that derivational paradigms in any language will be restricted and morpheme productivity limited, for some signed languages at least the degree of grammaticization of these modifications is as yet uncertain.

Morphemes and morphological processes can also signal the inflection of existing signs, adding grammatical information (e.g., marking for number, person, aspect, etc.) while maintaining the essential lexical meaning of the stem. Inflections in signed and spoken languages can be found on nominals or predicates (verbs and adjectives).

Inflection by concatenative affixation appears to be extremely rare in signed languages. One example is a nominal genitive suffix in Auslan which is used to signal possessive relationships between two nouns. This sign is not used as a free morpheme of any kind, but only as a suffix, as in MOTHER+GEN SISTER+GEN HUSBAND to mean 'mother's sister's husband.'

Data from a growing number of signed languages have shown that all of them exploit space and movement patterns inherent in sign formation to convey information regularly encoded in the inflectional systems of spoken languages. Indeed, the markings often appear to be in part phonologically conditioned.

For example, space will be exploited when a sign is not anchored throughout its production at a particular location on the body, or if its movement parameter is not specified for repetition.

These processes are exemplified in both nominal and verbal inflections. In nominal signs, inflection for plurality can be marked by repetition (often in different locations if the sign is not anchored). Spatial modifications can also signal topographical information about referents (e.g., a noun may be placed in the signing space to mean 'thing-at-such-and-such-a-place').

A basic tripartite division of ASL verb signs as plain, spatial, and agreeing (Padden, 1988) has been found to apply to signed languages to which the framework has been applied (with or without various modifications and reinterpretations).

For example, no sign language lacks the ability to modify the direction of the movement parameter of agreeing verbs so that the beginning and end points of these signs move between regions in the signing space associated with the *subject* (or the *agent*) and the *object* (or the *patient*) of the action. (Alternative interpretations of this phenomenon include those that assign *source* and *goal* as the underlying significance of these locations (Johnston, 1991; Meier, 1982).) The sign glossed as GIVE in many signed languages can have the meanings 'I give you' when moved from the signer to the interlocutor, or 'you give me' when moved from the interlocutor toward the signer, and 'he/she gives him/her' when moved between two (real or imaginary) third entities in the signing space. In spatial verbs the same mechanism is exploited to mark spatial and locative information. (Plain verbs are unable to exploit location and direction of movement in this way because they have a fixed place of articulation.) See Figure 7. For Padden and many linguists, the modifications on agreeing verbs are analyzed as non-concatenative affixes inflecting for person, while for other linguists the modifications indicate locations, depicting actions within a mental space representation of an event (Liddell, 2003).

A second related phenomenon in verb inflection refers to distribution and involves a 'plural sweep' in which the end point is moved in an arc through locations associated with referents or relocated and redirected at each in a series of repetitions, as in the modification of ASK to mean 'ask all' or 'ask each.' See Figure 8.

Similarly, modifications can be made to the movement parameter of verb signs to express a number of aspectual nuances. Researchers in many signed languages have identified similar patterns of movement modification, with similar meanings, as first described in ASL (Klima *et al.*, 1979). For example, similar patterns of cyclic and repeated movements also convey durational and continuative aspect (e.g., 'ask repeatedly'). Aspectual and distributional modifications can also combine to create a morphologically complex multilayered pattern of modifications (e.g., 'ask all repeatedly').

Verbal modifications based on suprasegmental modifications involve nonmanual features such as facial expression. A large number of facial expressions have been identified across many signed languages. Two found in Auslan/BSL – 'th' (as if producing an interdental fricative) and 'mm' (a bilabial protrusion) – are examples also found in some other signed languages. The former implies lack of control or inattention, the latter implies relaxed normality and when co-articulated with DRIVE they mean 'drive carelessly' and 'drive relaxed and normally', respectively.

As with the derivation of nominals, the extent of the grammaticization of movement modifications inflecting for aspect and manner, and their obligatoriness, within many signed languages, is still yet to be determined, or, at minimum, appears to vary.

Though many signs are monomorphemic or bimorphemic, highly iconic lexical signs, of which there are significant numbers in any signed language, often have more than two identifiable morphemes – they are multimorphemic. Take the highly iconic sign DRINK/CUP (as if holding a cup to one's mouth) which is found in many signed languages with a similar form



Figure 7 Inflection ('person agreement') through movement modification in the Auslan/BSL sign GIVE, and the Auslan sign HELP. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

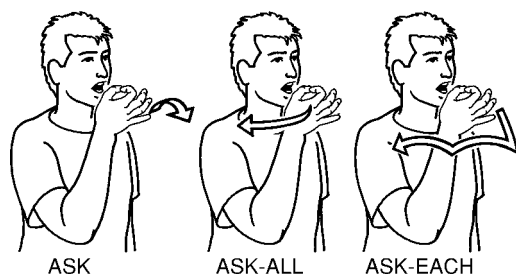


Figure 8 Inflection ('distribution') through movement modification in Auslan/BSL. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

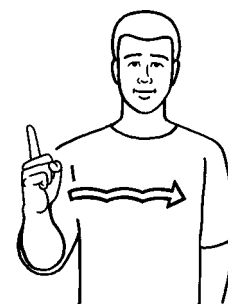
and meaning. It consists of at least three morphemes: The handshape signifies holding a cup, the movement signifies turning a cup toward the mouth, and the location signifies the mouth.

Importantly, there are a significant number of signs produced in many utterances in any signed language which are neither lexical nor grammatical signs. They display a moderate to high degree of conventionality in the form and meaning of handshapes, while movements and locations appear to be draw on particular representational exigencies of the moment. These signs are variously called classifier signs or polymorphemic signs.

Polymorphemic signs are used to convey a large amount of visual spatial information about participants in a situation (e.g., the size and shape and location of entities, how they may be handled, their position in space, and their path and manner of movement). In the following sign, the upright index finger represents a person, the palm the front of the body, the movement left to right the path, and the bobbing up and down movement a walking action. See Figure 9.

No signed language appears to lack these types of signs, and a considerable literature has been generated in an attempt to analyze them systematically in linguistic terms (Emmorey, 2003; Engberg-Pedersen, 1993; McDonald, 1982; Schick, 1990; Supalla, 1986). They can create monosyllabic polymorphemic signs which are unattested in spoken languages and which resist analysis into roots and listable morphemes. They remain a problem area for linguists (Schembri, 2003).

Original research into signed languages aimed to establish them as real languages with language-like characteristics. Subsequent research has aimed to establish the validity of linguistic universals that had been made on the basis of the study of spoken languages only, or to determine the impact of modality on language structure (e.g., Meier *et al.*, 2002).



PERSON-WALK-BY-FROM-RIGHT-TO-LEFT

Figure 9 A complex polymorphemic sign found in many signed languages. Reproduced from Johnston T & Schembri A (eds.) (2003). *The survival guide to Auslan: a beginner's pocket dictionary of Australian Sign Language*. Sydney: North Rocks Press with permission.

Another line of research seeks to acknowledge the degree to which signed languages are different from spoken languages. Depending on how the dynamics of spoken language are understood, these differences have been perceived as additional special characteristics peculiar to language in the visual-gestural modality, or as differences of degree only which have been occasioned by modality, e.g., some representational resources, such as space, are universally exploited in signed languages. For some linguists, the exploitation of a spatial and iconic morphology is seen as unique to signed languages but is nonetheless analyzed as part of a fully linguistic system of agreement (Aronoff *et al.*, 2000), for others it represents a fusion of elements of language and gesture (Liddell, 2003). For yet others, these face-to-face representational resources are recognized as available to *all* language users – even if they are underexploited in spoken languages and are ignored in most grammars. They saturate the lexico-grammar of signed languages because they are always available in languages that are embodied and, of necessity, always in view (Johnston, 1992).

It is as yet unclear if all of the phenomena of sign language morphology can be properly dealt with as 'linguistic,' narrowly defined. Insofar as it may contribute to the redefinition of what is 'language' or what is properly 'linguistic,' the short history of the study of signed languages belies its relative importance to linguistics.

Bibliography

- Aronoff M, Meir I & Sandler W (2000). 'Universal and particular aspects of sign language morphology.' *University of Maryland Working Papers in Linguistics* 10, 1–33.

- Bellugi U & Fischer S D (1972). 'A comparison of sign language and spoken language.' *Cognition* 1, 173–198.
- Emmorey K D (ed.) (2003). *Perspectives on classifier constructions in sign languages*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Engberg-Pedersen E (1993). *Space in Danish Sign Language: the semantics and morphosyntax of the use of space in a visual language*, vol. 19. Hamburg: Signum Press.
- Fernald T B & Napoli D J (2000). 'Exploitation of morphological possibilities in signed languages: comparison of American Sign Language and English.' *Sign Language & Linguistics* 3(1), 3–58.
- Johnson R E & Liddell S K (1986). 'ASL compound formation processes, lexicalization, and phonological remnants.' *Natural Language and Linguistic Theory* 4, 445–513.
- Johnston T (1991). 'Spatial syntax and spatial semantics in the inflection of signs for the marking of person and location in Auslan.' *International Journal of Sign Linguistics* 2(1), 29–62.
- Johnston T (1992). 'The realization of the linguistic meta-functions in a sign language.' *Language Sciences* 14(4), 317–353.
- Klima E S, Bellugi U, Battison R, Boyes-Braem P, Fischer S D, Frishberg N *et al.* (1979). *The signs of language*. Cambridge, MA: Harvard University Press.
- Liddell S K (1984). 'Think and believe: Sequentiality in American Sign Language.' *Language* 60, 372–399.
- Liddell S K (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.
- McDonald B (1982). *Aspects of the American Sign Language predicate system*. Ph.D. diss., State University of New York, Buffalo.
- Meier R P (1982). *Icons, analogues, and morphemes: the acquisition of verb agreement in American Sign Language*. Ph.D. diss., University of California, San Diego.
- Meier R P, Cormier K A & Quinto-Pozos D (2002). *Modality and structure in signed and spoken languages*. Cambridge: Cambridge University Press.
- Padden C (1988). *The interaction of morphology and syntax in American Sign Language*. New York: Garland.
- Sandler W (1995). 'One phonology or two? Sign language and phonological theory.' *Glott International* 1(3), 3–8.
- Schembri A (2003). 'Rethinking "classifiers" in signed languages.' In Emmorey K D (ed.) *Perspectives on classifier constructions in sign languages*. Mahwah, NJ: Lawrence Erlbaum Associates. 3–34.
- Schick B S (1990). 'Classifier predicates in American Sign Language.' *International Journal of Sign Linguistics* 1(1), 15–40.
- Supalla T (1986). 'The classifier system in American Sign Language.' In Craig C (ed.) *Noun classes and categorization*. Amsterdam: John Benjamins. 181–214.
- Supalla T & Newport E L (1978). 'How many seats in a chair? The derivation of nouns and verbs in American Sign Language.' In Siple P (ed.) *Understanding language through sign language research*. New York: Academic Press. 91–132.
- Sutton-Spence R & Woll B (1999). *The linguistics of British Sign Language: an introduction*. Cambridge, UK: Cambridge University Press.
- Wilbur R B (1993). *Syllables and segments: hold the movements and move the holds!* New York: Academic Press.
- Zeshan U (ed.) (2006). *Interrogative and negative constructions in sign languages*. Oxford: Oxford University Press.

Sign Languages of the World

U Zeshan, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The Current State of Knowledge

After more than 30 years of systematic sign language research, most sign languages throughout the world still remain scarcely documented or even entirely unknown. We can only estimate how many sign languages exist in the world, and we are even less sure about how they may be grouped into language families. A few sign languages in industrialized countries are reasonably well documented, whereas little is known about sign languages in other areas of the world, such as sub-Saharan Africa, Southeast Asia, and the Arab world. Nevertheless, increasingly more

information has been coming to light during the past decade, although we are still far away from systematic linguistic documentation in most cases.

Based on what we know to date, it is fairly clear that the sign languages of the world number in the hundreds rather than in the thousands and are thus much fewer in number than their spoken counterparts. For all we know, they are also much younger than spoken languages, although other forms of gestural communication are as old as humanity itself. The latest edition of the *Ethnologue* (Grimes, 2004) lists approximately 100 living sign languages. However, there are many omissions and errors in this list, so the actual number of sign languages in the world is likely to be at least three or four times greater.

The maximum documented age for a sign language is slightly more than 500 years for the sign language

used at the Ottoman court in Turkey (Miles, 2000). There is no reason why the large cities of antiquity more than 2000 years ago should not have had groups of sign language users, but we do not have any reliable sources for these times. On the other hand, it is quite unlikely that communities of sign language users as we know them today would have existed even earlier. Only after urbanization had created reasonably large populations could critical numbers of deaf people theoretically have come together to use a sign language.

For many known sign languages, there is more or less detailed anecdotal evidence of historical links with other sign languages. These links may have to do with colonial history, migration of populations, or, in more recent times, the establishment of deaf education with the help of another country. The principal difficulty lies in determining whether a particular relationship between sign languages is genetic in nature (i.e., in how far we can speak of a sign language family) or whether we are dealing with a language contact situation. Attempts at addressing this issue have been largely unsuccessful, and no theoretically sound method of investigating historical relationships between sign languages is available.

In recent years, increasingly more sign languages are beginning to be documented. A first step is usually the compilation of basic vocabulary in word lists (pairing a word and a picture of a sign), which are often wrongly called 'dictionaries' (see **Figure 1**). During the past decade, these and other developments have resulted in a situation in which it is now possible to systematically compare linguistic structures across a much wider range of sign languages than in the past. The newly emerging field of sign language typology is concerned with the issue of how to systematize this new knowledge in a theory of variation across sign languages.

Sociocultural and Sociolinguistic Variables

Signed communication occurs in a variety of situations. This article is concerned exclusively with natural full-fledged sign languages that are the primary languages of their users. We are not concerned with artificially created sign systems such as 'Manually Coded English,' 'Signed Japanese,' and 'Dutch in Signs,' which have been invented for educational purposes with the aim of mirroring spoken language structures 'on the hands'. We are also not concerned with secondary sign languages that are used in communities where the usual mode of communication is through a spoken language but where signed communication plays a supplementary role for certain

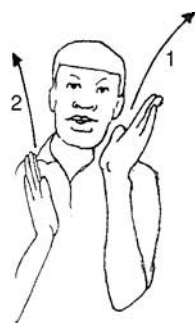
purposes, such as conditions of speech taboo. Rather, the sign languages we are interested in involve groups of deaf people for whom the sign language is the primary means of communication.

The first sign languages that were documented in detail from the 1970s onwards are used by communities of deaf people in urban settings. These are minority languages in which most of the users are deaf and there is constant language contact with the surrounding spoken/written language of the majority culture of hearing people. This situation is well described and occurs in urban areas in all regions of the world.

However, sign languages also exist in an entirely different sociocultural setting that is less well documented but highly significant for cross-linguistic comparison. These sign languages are used in village communities with a high incidence of hereditary deafness. Village-based sign languages arise because deaf individuals have been born into the village community over several generations, and therefore a sign language has evolved that is restricted to the particular village or group of villages. These sign languages are typically used by the whole village population no matter whether deaf or hearing, and in this sense, they are not minority languages, nor do they face any linguistic oppression. They have developed in isolation from other sign languages and are not used in any educational or official context. Deaf people are fully integrated into village life and may not be considered to be 'disabled' in any sense (Branson *et al.*, 1999). The existence of village-based sign languages has been reported from places as diverse as Bali, Ghana, Thailand, Mexico, an Arab Bedouin tribe in Israel, and a native Indian tribe in the Amazon, but their linguistic documentation is only just beginning. These languages have the potential to call into question many of the general assumptions that were made previously about the structure of sign languages.

Some village-based sign languages are already endangered and have not been documented in detail. As the larger, urban sign languages move in through formal education and the media, these small, locally restricted sign languages face similar pressures as their spoken language counterparts (see **Endangered Languages**). Similarly, sign languages in some developing countries have been under pressure from foreign sign languages, as in many African countries. In places where the deaf community is very large and the indigenous sign language has had time to develop on its own, it is relatively immune to foreign influences, as is the case in China and in the Indian subcontinent.

Despite similarities with respect to language endangerment, the life cycle of sign languages also differs from that of spoken languages in that new sign languages continuously emerge throughout the world,



NANASI
pineapple

Hutumika pia kwa Katani
it is also used for "sisal"

169:571

Alama sanifu
standard sign



NAZI
coconut



Alama sanifu
standard sign

LANGUAGE RULES AND TERMINOLOGY OPPOSITES

قواعد الفاظ وتران متضاد الفاظ



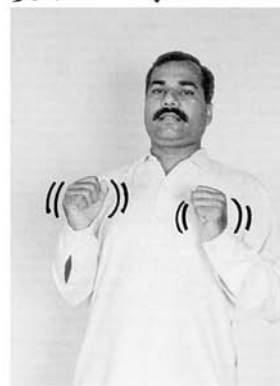
LESS

کم



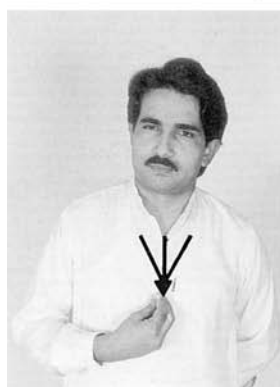
MORE

زیادہ



COLD

سرد



SAD

اُداس



HAPPY

خوش



HOT

گرم

Figure 1 Entries from sign language dictionaries (Tanzania, Pakistan).

as most famously documented in Nicaragua (Kegl *et al.*, 1999). Throughout the world, urbanization and the spread of special education for the deaf create new deaf communities with newly emerging sign languages. The stage of a sign language's life cycle is an important consideration for comparing the structures of sign languages.

Relationships between Sign Languages

For a number of individual sign languages as well as groups of sign languages, the notion of sign language family has been proposed, based on known facts about their relationship with each other. For example, it is well-known that sign language was brought to New Zealand and Australia from the United Kingdom, and therefore these three sign languages make up the 'British Sign Language family.' For different historical reasons, the Japanese Sign Language family includes sign languages in Taiwan and Korea, both of which had been under Japanese occupation. In cases in which one and the same sign language-using community seems to have split and subsequently developed independently from each other, the traditional family tree model can be applied, and the shared history is visible and interpretable. Sign languages in Australia, New Zealand, and the United Kingdom are still mutually intelligible to a large extent and share most of their vocabulary, to the extent that it is doubtful whether they should not be classified as dialects of one and the same language. Sign languages in Korea, Japan, and Taiwan all share a peculiar grammatical mechanism of gender marking, with the thumb indicating male and the little finger female gender as formative elements in complex signs (see Figure 2). This feature is not found in any other known sign language and, together with other factors, makes a strong case for positing a shared history of this sign language family.

However, the situation is usually not so clear-cut. In most cases, it is impossible to determine whether similarities between two sign languages are the result of a genetic relationship or the result of language contact. Instead of the 'pure' kind of family tree relationship, a more common type of relationship between two sign languages involves various kinds of language contact situations, language mixing, and creolization. For example, American Sign Language is said to have arisen in a creolization process, where Old French Sign Language came in contact with indigenous sign varieties, resulting in a new language with input from both of these sources. This kind of relationship cannot be considered genetic in the usual sense of the term.

In many cases, there is more or less clear historical evidence of relationships between sign languages. This may be related to colonial history so that, for instance, sign language communities in the Indian subcontinent use a two-handed manual alphabet as in British Sign Language. However, actual historical documentation of how this came to be the case is lacking, there are very few meaningful similarities in the vocabulary and grammar of the two sign languages, and there is thus no evidence for including Indo-Pakistani Sign Language in the British Sign Language family. Another common factor in linking two sign languages often involves the establishment of educational facilities for the deaf. For instance, the sign language in Brazil is said to have its root in French Sign Language because a deaf Frenchman established the first school for the deaf in Brazil, and Swedish Sign Language was similarly brought to Finland. We find this kind of link between many African countries and one or more Western 'source' sign languages (Schmaling, 2001). American Sign Language (ASL) has had a major impact on deaf communities in other countries, such as Thailand, the Philippines, Uganda, Zambia, Ghana, Malaysia, and Singapore,



Figure 2 Gender marking in South Korean Sign Language: SCOLD(someone), SCOLD(me), SCOLD(a male person), SCOLD(a female person).

and it is often unclear whether the sign languages used in these countries should be considered dialects of ASL, descendants of ASL in a family tree of languages, ASL-based creoles, or independent sign languages with extensive lexical borrowing from ASL. To the extent that indigenous sign languages already existed in these countries and secondarily came under the influence of a foreign sign language, the relationships between them are not genetic in the usual sense but are instances of language contact.

This kind of problem is not unknown for spoken languages but is aggravated by a number of complicating factors in the case of sign languages. First, the familiar historical-comparative method that is used to determine language families and reconstruct older forms of source languages has never been applied to sign languages. No process of regular sound change has been identified, and the comparison of morphological paradigms is often compromised because the forms in question are iconically motivated. Vocabulary comparisons are highly unreliable, and there seems to be a considerable ‘baseline level’ of iconically determined lexical similarity even between unrelated sign languages (Guerra Currie *et al.*, 2002). The first family trees that were proposed for sign languages were based on historical evidence and lexical similarities, and later attempts at using glottochronology on the basis of word list comparisons (Woodward, 1993, 2000) are similarly unreliable.

Another complicating factor in many cases is the uncertainty about whether or not there were indigenous sign varieties before the influence of a foreign sign language set in and, if so, what the linguistic status of this signed communication might have been. It is possible that in a particular region, limited home sign systems came in contact with a foreign full-fledged sign language, resulting in a new sign language in a process that has no counterpart among spoken languages. Finally, the lack of any historical records makes it difficult to directly test and evaluate any proposed historical relationship between sign languages. In the absence of any sound methodology for establishing sign language families, the issue of how one sign language is related to another one usually remains unresolved.

Grammatical Similarities and Differences across Sign Languages

Over time, sign language linguists have come to expect certain features in the structure of sign languages that have been shown to occur with great regularity in most or all sign languages known and described so far. Accordingly, there are attempts at accounting for these putative sign language universals on the basis of

their visual-gestural modality. For instance, sign languages offer the possibility of using spatial grammatical mechanisms by virtue of being three-dimensional languages, and therefore they tend to use movement modifications to express aspectual distinctions or to use movement direction to code verb agreement. Since the articulators in sign language are larger and slower than in a spoken language, sign languages tend to mark grammatical functions in a simultaneous rather than a sequential fashion; therefore, they use nonmanual behaviors such as facial expressions to mark sentence types (questions, negation, and subordination), and they use complex signs with numeral incorporation (e.g., a single complex sign meaning ‘three months’) (*see Sign Language: Morphology*). It has been claimed that sign languages are similar in the kinds of complex simultaneous morphology just mentioned but differ from each other in sequential morphology such as clitics and affixes, with sequential morphology being comparatively rare in sign languages (Aronoff *et al.*, 2000).

Most of these generalizations about the similarities between sign languages are based on investigations of a limited number of languages, mainly in Europe and North America. The picture changes somewhat when examining a larger range of the world’s sign languages. Although the previous observations are indeed true of many sign languages throughout the world, this is only part of the story. First, some sign languages do not show the ‘expected’ types of structures. Two unrelated village-based sign languages, in Bali and Israel, do not show an elaborate system of spatial verb agreement as is familiar from other sign languages. Another village-based sign language in Ghana does have spatial verb agreement but does not use the so-called ‘classifier’ hand shapes to refer to categories of moving persons, animals, and vehicles. Given that village-based sign languages have developed in isolation from any other sign language and exist under very different sociolinguistic conditions, it is not unexpected to find important differences in their structures in comparison with urban sign languages.

The range of possible structures in sign languages expands considerably when we consider non-Western, lesser-known sign languages. The gender marking system in the Japanese Sign Language family represents one such example. Sign language varieties in China also show many particularities that are not familiar from documented Western sign languages. Chinese Sign Language varieties include so-called ‘character signs,’ a particular type of borrowing in which the shapes and movements of the hands imitate the whole or part of words from the Chinese writing system (*see Figure 3*). Both northern and southern

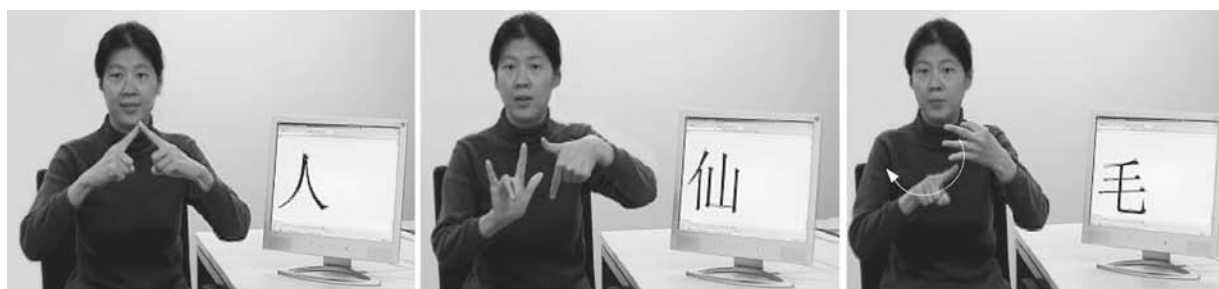


Figure 3 Character signs in Chinese Sign Language.



Figure 4 Chinese Sign Language signs with “little finger” negative morpheme: DEAF and TASTELESS.

sign language varieties in China also make use of a productive mechanism of negation in which negative signs are marked by an extended little finger and the positive counterparts have an extended thumb (see Figure 4). Finally, question words for quantifiable concepts include one or two open hands with finger wiggling as part of complex signs, forming a large paradigm of interrogatives. The study of a greater range of sign languages thus reveals a large number of previously undocumented grammatical structures, just as the study of ‘exotic’ spoken languages did in earlier stages of spoken language linguistics.

Other structural differences between sign languages are more subtle and only come to light after systematic investigation. Typologically oriented studies across sign languages exist for a limited number of grammatical domains (for pronouns, see McBurney, 2002; for questions and negation, see Zeshan, 2004a, 2004b). Such studies show that the degree of structural differences between sign languages may be considerable but is unevenly distributed across different parameters of investigation. For example, sign languages differ as radically as spoken languages with respect to the set of their possible question words. A sign language may have only a single question word, as in certain dialects of Indo-Pakistani Sign

Language (see Figure 5), or more than a dozen, as in Hong Kong Sign Language. On the other hand, the facial expressions accompanying questions tend to be very similar across unrelated sign languages, with eye contact, forward head position, and eyebrow movement as prominent features. Understanding the reasons for these patterns is important for building a theory of typological variation across sign languages.

Another important result from comparative studies is that certain sign language forms may look very similar superficially but in fact have very different properties. For instance, in a broad range of 38 sign languages throughout the world (see Figure 6), it has been found that in each case, negation can be expressed by a side-to-side headshake (Zeshan, 2004a). However, the grammatical constraints governing the use of headshake negation in fact differ greatly across sign languages. Whereas in some sign languages, such as in the Scandinavian region, headshake negation is a primary negation strategy and may often be the only instance of negation in the clause (Bergman, 1995), other sign languages, such as in Japan and Turkey, obligatorily use a manual negative sign with or without headshake negation as a secondary accompaniment. Sign languages in the eastern Mediterranean region (Greece, Turkey, and neighboring Arab countries) additionally use a single backward head tilt for negation that has not been found in any other region of the world (Zeshan, 2002).

It can be assumed that the significance of many possible parameters of variation across sign languages has not been recognized. For example, mouth movements deriving from a silent representation of spoken words, so-called ‘mouthing,’ carry an important functional load in some sign languages (e.g., in Germany, The Netherlands, and Israel) but are functionally largely irrelevant in Indo-Pakistani Sign Language (Boyes Braem and Sutton-Spence, 2001). The presence or absence of contact with literacy may be another important factor, evidenced by the fact that not all sign languages use an indigenous manual alphabet for fingerspelling.



Figure 5 Combinations with the Indo-Pakistani Sign Language question word (WH): PLACE+WH “where”, TIME+WH “when”.



Figure 6 Sign languages represented in the typological survey on questions and negation.

Future Developments

The dynamics of developments throughout the world with respect to sign languages and their documentation carry considerable momentum. Some sign languages are endangered, whereas others are expanding in geographical spread and contexts of use, and some are only just being created by new communities of users. Forces such as intensive contact between sign language and spoken language, as well as between one sign language and another, and the move toward official recognition for sign languages and the deaf communities that use them rapidly change and reshape the makeup of many sign languages worldwide. It is a continuous challenge

for sign language linguistics to keep up with these developments and put together an increasingly detailed picture of linguistic diversity among the world's sign languages.

Bibliography

- Aronoff M, Meir I & Sandler W (2000). ‘Universal and particular aspects of sign language morphology.’ In Grohmann K K & Struijke C (eds.) *University of Maryland working papers in linguistics*, vol. 10. College Park: University of Maryland Press. 1–34.
- Baker A, van den Bogaerde B & Crasborn O (2003). *Cross-linguistic perspectives in sign language research. Selected papers from TISLR 2000*. Hamburg: Signum.

- Bergman B (1995). 'Manual and nonmanual expression of negation in Swedish Sign Language.' In Bos H F & Schermer G M (eds.) *Sign language research 1994. Proceedings of the fourth European Congress in Sign Language Research, Munich, September 1994*. Hamburg: Signum. 85–114.
- Boyes Braem P & Sutton-Spence R (eds.) (2001). *The hand is the head of the mouth: the mouth as articulator in sign languages*. Hamburg: Signum.
- Branson J, Miller D & Marsaja I G (1999). 'Sign languages as a natural part of the linguistic mosaic: the impact of deaf people on discourse forms in North Bali, Indonesia.' In Winston E (ed.) *Storytelling and conversation. Discourse in deaf communities*. Washington, DC: Gallaudet University Press. 109–148.
- Erting C J, Johnson R C, Smith D L & Snider B D (eds.) (1994). *The deaf way. Perspectives from the International Conference on Deaf Culture*. Washington, DC: Gallaudet University Press.
- Grimes B F (ed.) (2004). *Ethnologue: languages of the world* (14th edn.). Dallas, TX: Summer Institute of Linguistics.
- GuerraCurrie A-M, Meier R & Walters K (2002). 'A cross-linguistic examination of the lexicons of four signed languages.' In Meier *et al.* (eds.). 224–239.
- Kegl J, Senghas A & Coppola M (1999). 'Creation through contact: sign language emergence and sign language change in Nicaragua.' In DeGraff M (ed.) *Comparative grammatical change: the intersection of language acquisition, creole genesis, and diachronic syntax*. Cambridge: MIT Press.
- McBurney S (2002). 'Pronominal reference in signed and spoken language: are grammatical categories modality-dependent?' In Meier *et al.* (eds.). 329–369.
- Meier R, Cormier K & Quinto-Pozos D (eds.) (2002). *Modality and structure in signed and spoken languages*. Cambridge, UK: Cambridge University Press.
- Miles M (2000). 'Signing in the Seraglio: mutes, dwarfs and jestures at the Ottoman Court 1500–1700.' *Disability & Society* 15(1), 115–134.
- Schmaling C (2001). 'ASL in northern Nigeria: will Hausa Sign Language survive?' In Dively V L, Metzger M, Taub S & Baer A M (eds.) *Signed languages: discoveries from international research*. Washington, DC: Gallaudet University Press.
- Van Cleve J (ed.) (1987). *Gallaudet encyclopedia of deaf people and deafness* (3 vols). New York: MacGraw-Hill.
- Woodward J C (1993). 'The relationship of sign language varieties in India, Pakistan, & Nepal.' *Sign Language Studies* 78, 15–22.
- Woodward J C (2000). 'Sign languages and sign language families in Thailand and Viet Nam.' In Emmorey K & Lane H (eds.) *The signs of language revisited: an anthology to honor Ursula Bellugi and Edward Klima*. Mahwah, NJ: Erlbaum.
- Wrigley O *et al.* (eds.) (1990). *The Thai Sign Language dictionary* (Rev. and expanded edn.). Bangkok: NADT.
- Zeshan U (2002). 'Sign language in Turkey: the story of a hidden language.' *Turkic Languages* 6(2), 229–274.
- Zeshan U (2004a). 'Hand, head and face – negative constructions in sign languages.' *Linguistic Typology* 8(1), 1–58.
- Zeshan U (2004b). 'Interrogative constructions in sign languages – cross-linguistic perspectives.' *Language* 80(1), 7–39.
- Zeshan U (2005). 'Sign languages.' In Dryer M, Gil D & Haspelmath M (eds.) *World atlas of language structures*. Oxford: Oxford University Press.

Sindhi

J Cole, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Sindhi is an Indo-Aryan language with its roots in the lower Indus River valley. It takes its name from the Indus River, known in earlier times as the Sindhu. Today Sindhi is spoken in the province of Sindh, Pakistan, where it is recognized by the government as the official language of the province, home to an estimated 30–40 million people (projected from 1981 census data). Nearly half of the population of Sindh lives in rural areas, where Sindhi is the primary language. In the urban centers of Sindh, Sindhi competes for status and speakers with Urdu (the national language of Pakistan) and, increasingly, English. Sindhi

is also spoken by about 2.5 million people in India, including major communities in Gujarat, Mumbai, and Pune, where immigrants from Sindh relocated after the 1947 partition of India and Pakistan. Beyond the Indian subcontinent, Sindhi is spoken by large diaspora communities in the United Kingdom and the United States, and around the world.

Language History

Sindh is the site of the ancient Harappan civilization of the lower Indus River valley. A case can be made that remnants of Harappan culture are evident in classical Sindhi folklore and religious rituals, which raises the question of a possible linguistic link between Sindhi and the Harappan language. Unfortunately, there is little evidence on which to determine the

linguistic stock of the Harappan language, the ancient script is as yet undeciphered, but a prevailing theory suggests a Dravidian origin. This theory points to the presence of the Dravidian language Brahui, spoken in the northwestern Pakistani province of Baluchistan, as a remnant of a broader Dravidian region in the subcontinent in earlier times. This possible link to an ancient Dravidian language of the Harappans has led some scholars to claim a Dravidian origin for Sindhi. This minority view, however, clashes with a substantial body of linguistic evidence for an Indo-Aryan origin of Sindhi.

The earliest historical reference to Sindhi is in the *Nātyaśāstra*, a dramaturgical text that was written between 200 B.C. and 200 A.D. Evidence for Sindhi as a written language dates to a translation of the Islamic *Qur'an* in 883 A.D., followed a century later by a Persian translation of the ancient Indian religious epic *Mahābhārata* taken from a language thought to be Old Sindhi. Dating the emergence of Sindhi in the evolution of Indo-Aryan is a matter of some controversy. Various theories, ably summarized by Khubchandani (2000), trace Sindhi to the Vṛāṇa Apabhramśa or to an earlier pre-Vedic Prakrit language. Although Trumpp (1872), in his authoritative Sindhi grammar, describes Sindhi as a more 'pure Sanskritical' language compared to the other modern Indo-Aryan languages, Sindhi undeniably reveals the impact of its long history of contact with speakers of other languages.

Sindh has succumbed to foreign rule many times over a history of 2500 years and, much like English, has accumulated linguistic features and vocabulary from the languages of its foreign rulers. In pre-Muslim history, from the 6th century B.C. through the 1st century A.D., Sindh was invaded by a succession of Achaemenian, Greek, Mouryan, Scythian, and Persian rulers, including Alexander the Great (329–324 B.C.). After a brief period of rule by local dynasties, the Arab invasion in 711 A.D. initiated the Muslim period and the heavy influence of Persian on Sindhi with numerous lexical borrowings. Following a period of rule by local dynasties from the 11th through the mid-19th century, Sindh joined the British Empire in 1843. The influence of English on Sindhi, especially through lexical borrowings, began at that time and continues in the present, and is second only to Persian (Western Farsi) in its impact on the language. The result of language contact in all these periods of foreign rule is a Sindhi lexicon with diverse etymological bases and multiple cognate forms, which is further complicated by an exceptional number of irregular verbal inflections, and by the expansion of the sound inventory to include several Perso-Arabic sounds not native to Indo-Aryan.

Related Languages and Dialects

Among the languages spoken in the region today, Sindhi is closely related to Sirāikī (Saraiki), spoken north of Sindh province, and to Kachchhī (Kachchi), spoken in the Kachchh region of Gujarat, along the border between Pakistan and India. Grierson's (1919) survey listed five regionally defined Sindhi dialects, with Vicholī ('Central') as the standard variety. Contemporary dialectal work has been carried out by Khubchandani (1962–1963) and Rohra (1971) on Kachchhī (also spelled Kachhi, Kachchhi). Bughio's (2001) sociolinguistic study of the urban and rural Vicholī varieties is the only work since Grierson to deal with Sindhi dialect variation within Pakistan, and opened the door for promising future investigation.

Linguistic Features

Sound System

Sindhi and the neighboring languages Sirāikī and Marwari are distinct among Indo-Aryan languages for their use of the glottal implosive stops /ɓ, ɗ, ɟ/, which derive from Middle Indo-Aryan geminate voiced stops in medial position and single voiced stops in initial position. In other respects, the sound inventory of Sindhi is typical of Indo-Aryan, with a full series of voiceless, voiced, aspirated, and voiced aspirated stops and nasal stops at five places of articulation (see Table 1). Alongside the alveolar rhotic tap [ɾ] there is a retroflex tap [ɽ]; but unlike [ɾ], the retroflex tap is restricted to intervocalic position, where it can be considered the positional variant of the retroflex stop [ɖ]. Retroflex [ɖ] occurs intervocalically only in a few English loan words, where it corresponds to the English alveolar [d], as in *loḍḍinga* 'truck' and *reḍḍiyo* 'radio' (from English *loading*, *radio*). Sindhi has incorporated a number of consonants from Persian, including the well-established sounds /f, v, ʃ, z, ʒ/, along with /q, ɣ/, which are not typically used except by urban, educated speakers for whom they are arguably reinforced by their stable presence in Urdu.

Sindhi has the standard Indo-Aryan vowel inventory with ten vowels that can be grouped in five long–short pairs: /i, iː, e, eː, u, uː, o, oː, a, aː/. The short mid vowels are subject to dialectal variation or merger (discussed below). Long vowels can occur with contrastive nasalization; compare the final long nasal vowels in *sa:mboː* 'in front of' and *vitʰuː* 'scorpion', with the final long oral vowels in *sanhoː* 'thin' and *kaduː* 'gourd'. Phonetically nasal short vowels occur in the context of a following tautosyllabic nasal consonant, e.g., *āmbu* 'mango', but can also occur in an open syllable preceding or following /h/, where they

Table 1 The consonants of Sindhi

| | Labial | Dental | Alveolar | Post-alveolar | Palato-alveolar | Velar | Glottal |
|-----------|--------------------------------------|--------------------------------------|----------|--|--|--------------------------------------|---------|
| Stop | p b p ^h b ^h | t d t ^h d ^h | | t̪ d̪ t̪ ^h d̪ ^h | | k g k ^h g ^h | |
| Implosive | ɓ | | ɗ | | ɟ | ɣ | |
| Affricate | | | | | tʃ dʒ tʃ ^h dʒ ^h | | |
| Nasal | m | | n | ɳ | ɲ | ŋ | |
| Fricative | f | | s (z) | | (ʃ) | (x ɣ) | h |
| Rhotic | | | r | (ɽ) t ^h | | | |
| Lateral | | l | | | | | |
| Glide | ʋ | | | | y | | |

contrast with oral short vowels. Compare the short nasal vowels in *mūhū* ‘mouth’, *jīhī* ‘who/relative pronoun’ with the short oral vowels flanking /h/ in *māhalu* ‘palace’, *subahu* ‘morning’.

Sindhi syllable structure allows for at most one consonant to appear in the onset and coda position (CVC). Consonant clusters (CC) occur word medially, as in CVC.CV *kursi*: ‘chair’. With a few loanword exceptions, Sindhi words must end in a vowel. Short vowels in word-final position are extremely reduced, though grammatically important as markers of noun number, gender and case. Vowel-initial syllables may occur initially and medially, and in the latter case may give rise to word-internal vowel sequences (hiatus), as in *b^ha:fai:* ‘brother’s wife’. Hiatus sequences never occur with identical vowels.

There are several features of Sindhi pronunciation that are subject to dialectal variation, distinguishing the speech of rural, uneducated, or older speakers who represent an older variety of Sindhi, from urban, educated (i.e., literate) or younger speakers, whose speech is more noticeably influenced by Urdu, Hindi, and English pronunciation patterns. Three dialectal features described by Bughio (2001) are as follows. The short vowels /e, o/ are typically merged with their long counterparts in the old variety, resulting in [e, o] while new variety speakers more frequently keep them distinct, producing long monophthongs [e, o] and short diphthongs [ai, au] or lax vowels [ɛ, ɔ]. The diphthong realization is typical of Muslim new variety speakers, and the lax vowels are typical of Hindu new variety speakers. This distinction based on religious affiliation reflects in part the separation of Hindu and Muslim communities since the 1947 partition of India and Pakistan, and the maintenance of diphthongs in Arabic loan words (borrowed through Persian) in the speech of both Sindhi- and Urdu-speaking Muslims. Old and new varieties of Sindhi are also distinguished by the frequent deletion or total loss of the word-final short vowels in the new varieties. The third dialectal feature in

Sindhi is the pronunciation of the retroflex stops /t̪, d̪, d̪^h/ as stop-rhotic clusters [t̪r, d̪r, d̪^hr] in the old varieties.

Morphology

Sindhi has a rich system of nominal and verbal morphology, with regular paradigms of declension and conjugation that exist alongside a remarkably high number of exceptional forms. Nouns, adjectives, and pronouns are marked for number, gender, and case. The gender class of the noun is in most cases marked by the final vowel, and number and case marking are expressed through a combination of stem alteration and final vowel suffix. Examples of nominal declension are shown with paradigms for the masculine noun ‘boy’ and feminine noun ‘table’ in **Table 2**.

Cases other than the ones shown in **Table 2** are marked through the use of a postposition following the noun in the oblique singular form, for example *g^hara k^he* ‘to the house’ (dative), *g^hara k^hā:* ‘from the house’ (ablative), *g^hara sā:* ‘with the house’ (comitative), *g^hara mē:* ‘in the house’ (locative), *g^hara jo:* ‘of the house’ (genitive). The genitive postposition is unique in that it is declined like an adjective, agreeing with the possessed noun in number, gender and case, as in (1):

- (1) t̪^hokire ja:
 boy.MASC.sing.OBL GEN.MASC.pl.NOM
 kita:ba
 book.MASC.pl.NOM
 ‘the boy’s books’
 t̪^hokire ji:
 boy.MASC.sing.OBL GEN.FEM.sing.NOM
 bīli:
 cat.FEM.sing.NOM
 ‘the boy’s cat’

Sindhi verbs are marked for aspect, tense, mood, and concordance (gender and number) through a complex system of modification of the verb stem, which may in addition be followed by a modal and

Table 2 Nominal declensions for masculine and feminine nouns

| | | Nominative | Oblique | Ablative | Vocative |
|---------|----------|------------------------|-------------------------|---------------------------|------------------------|
| 'boy' | Singular | tj ^h okiro: | tj ^h okire: | tj ^h okirā: | tj ^h okira: |
| | Plural | tj ^h okira: | tj ^h okirani | tj ^h okiraniā: | tj ^h okira: |
| 'table' | Singular | meza | meza | mezā: | meza |
| | Plural | mezū: | mezuni | mezuniā: | mezū: |

Table 3 Example verb forms based on the root *lik^h* 'write'

| | | |
|---------------------------|-------------------------------------|-----------------------------|
| Present unspecified | lik ^h e t ^h o | 'he writes' |
| Definite future | lik ^h ando | 'he will write' |
| Present habitual | lik ^h ando a:he | 'he writes (habitually)' |
| Present continuous | lik ^h i: rahiyo a:he | 'he is writing' |
| Unspecified perfective | lik ^h iyō | 'he wrote' |
| Subjunctive perfective | lik ^h iyō huje | 'he may have written' |
| Imperative | lik ^h u | 'Write!' (familiar) |
| | lik ^h o | 'Write!' (polite) |

an auxiliary verb. These elements combine in various ways to produce 17 distinct finite verb forms, and six nonfinite verb forms that function as nominal, adjectival, and adverbial participles. Each finite and nonfinite verb form can undergo further modification of the verb stem to express voice (active/passive) and valence (transitive/causative) distinctions. Several finite forms of the verb *lik^hanu* 'to write', all expressing masculine, singular concordance, are illustrated in Table 3.

Syntax

The pragmatically neutral word order in Sindhi is Subject-Object-Verb, but the order of these major constituents can be changed to put a phrase with Topic focus at the front of the sentence. Within phrases, the head element always occurs at the end, as in the noun phrase and verb phrase examples in (2):

- (2) *hiza nand^hri: tōpi:*
 this small hat
 'this small hat'
 ama: k^he tji^hi: lik^hi:
 mother DAT letter wrote
 'wrote a letter to mother'

The verb typically agrees with a nominative case-marked subject, as in

- huza atje t^hi:*
 she-NOM come-3sing aux.FEM
 'she comes'.

An 'experiencer' subject of a verb expressing physical state, psychological state, or kinship is marked with the dative postposition, as in

- huna k^he 6uk^hi laḡi:*
 he-OBL. DAT hunger-FEM strikes-FEM.sing.
 'he is hungry'.

Sindhi has the split-ergative agreement pattern found in other Indo-Aryan languages, whereby in the perfective aspect the subject of a transitive verb is marked for oblique case, and the verb agrees with a nominative (inanimate) object if present, and otherwise displays a default agreement (3sing.MASC).

Linguistic Works on Sindhi

Among published grammatical works on Sindhi, there are several grammars in the Sanskritic tradition, including Stack (1849), Trumpp (1872), and a section in Grierson's *Linguistic survey of India* (1919). Contemporary linguistic studies include instrumental phonetic studies (Nihalani, 1986, 1995), sociolinguistic and dialect studies (Rohra, 1971; Bhugio, 2001), and contemporary grammatical analysis (Khubchandani, 1961). Khubchandani (2000) presents a comprehensive bibliography of works on Sindhi from 1947–1967.

Bibliography

- Bhugio Q (2001). *A comparative sociolinguistic study of rural and urban Sindhi*. Munich: Lincom Europa.
 Grierson Sir G A (1919). *Linguistic survey of India* 7(1): *Indo-Aryan family: northwestern group, specimens of Sindhi and Lahnda*. Calcutta: Superintendent Government Printing.
 Khubchandani L M (1961). 'The phonology and morpho-phonemics of Sindhi.' M.A. thesis, University of Pennsylvania.
 Khubchandani L M (1962–1963). *The acculturation of Indian Sindhi to Hindi: A study of language in contact*. Ph.D. Dissertation, University of Pennsylvania (1962). Dissertation microfilm (1963). Reprinted in *Linguistics: An International Review*, vol. 12 (March 1965).

- Khubchandani L M (2000). *Sindhi studies 1947–1967: a review of Sindhi language and society*. Pune, India: Centre for Communication Studies.
- Masica C P (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Nihalani P (1986). 'Phonetic implementation of implosives.' *Language and Speech* 29, 253–262.
- Nihalani P (1995). 'Sindhi.' *Journal of the IPA* 25(2), 95–98.
- Panhwar M H (1983). *Chronological dictionary of Sind*. Jamshoro, Pakistan: Institute of Sindhology, University of Sind.
- Rohra S K (1971). 'Sindhi, Kacchi, and emigrant Sindhi.' *Indian Linguistics* 32(2), 123–131.
- Stack Capt G (1849). *Dictionary, English and Sindhi*. Bombay: American Mission Press.
- Trumpp E (1872). *Grammar of the Sindhi language*. London: Trübner and Co./Leipzig: F. A. Brockhaus.

Sinhala

J W Gair, Cornell University, New York, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

General

Sinhala (Sinhalese, Singhalese) is the first language of the majority in Sri Lanka, spoken by approximately 75% of the island's population (approximately 20 million in 2004) and thus has about 15 million first language speakers. It was declared the only official language of the country in 1956, but the status of official language was extended to the main minority language, Tamil, in 1987, with English as a link language. Sinhala belongs to the Indo-Aryan family and is thus related to languages of Northern India such as Hindi-Urdu, Panjabi, Bengali, and Marathi, and ultimately through Indo-European to English and the major Western European languages. Within Indo-Aryan, its closest relative is Dhivehi (Maldivian) of the Maldivian islands (also on Minicoy, where it is known as Mahl) with which it forms a southern sub-group. These two languages have been isolated from their sister North Indian languages for over two millennia and have developed special characteristics of their own, many of them shared, though the languages are mutually unintelligible. They have also been significantly influenced by the neighboring Dravidian languages, particularly Tamil-Malayalam. One common Sinhalese Buddhist tradition states that the language was brought to the island from northeast India, on the date of the final passing of the Buddha (544–543 B.C.E. in that tradition) though there are competing accounts. In any event, there are Sinhala inscriptions from the third and second centuries B.C.E. already showing changes distinguishing it from its sisters in India, most notably the complete loss of the aspirated consonant series, so that a date around the middle of the first millennium or shortly thereafter is not unreasonable. The place of origin has been disputed by serious scholars, one problem

being that the time of separation predates the major changes that differentiate the Indian Indo-Aryan dialects. Recent work on historical phonology, however, does give some evidence for a generally eastern origin, along with dialect admixture.

The extant literature in Sinhala dates from the 9th–10th centuries, but it is clear from Pali texts and later references that there was an earlier body of works that was lost due to internal turbulence. Fortunately, there is an epigraphical record from the 3rd–2nd centuries B.C.E. on, so that we have a continuous record of the language from that time, an asset unmatched in any other modern Indo-Aryan language. The literary legacy is rich. Not surprisingly, given the cultural history of the island, much of it is Buddhist in content, encompassing both prose and poetry, but there is a flourishing current literary scene in virtually all genres, including poetry, novels, short stories, criticism, and a lively theatre. Literacy is high, approximately 89%, and near 100% in some areas, which contributes to this productivity.

Sinhala vocabulary includes words from many languages, including English, Dutch, Portuguese, and others as well as Dravidian. Sanskrit loans abound, especially in the technical and religious/philosophical spheres, but perhaps unexpectedly, given the prevalence of Theravada Buddhism, Pali loans are scarce.

Varieties

While there is some regional dialect variation, all spoken dialects are mutually intelligible, requiring only some small adjustment. The main varietal difference is diglossia, involving two main functional varieties, generally referred to as Spoken Sinhala and Literary Sinhala. The former is used for all face-to-face communication, and the latter for virtually all written and published materials, except in some informal communication and in dialogue in modern fiction and drama, with the embedding narrative material generally in Literary. One major difference that

has been taken as the defining characteristic of the varieties is that Literary Sinhala has subject-verb agreement, as in (1), while all of the Spoken varieties lack it, as in (2). There are also accompanying differences in morphology, case forms, and use, and in fact the varieties differ on all levels of structure except phonology, since oral presentations of Literary use the spoken repertory. Literary productions are virtually always written beforehand. There are subvarieties within each of the major ones, and a Formal Spoken variety lacking agreement but showing some Literary features such as lexicon and some case features is seeing increasing use, partly because of the need for live production in the media, as in interviews and speeches. Some news broadcasts now use that variety, though earlier they were written in Literary and read out.

(1) Literary (in transliteration):

mama goyam kapami.
I rice-plants cut-PRESENT-1SG
'I reap paddy.'

api goyam kapamu.
We rice-plants cut-PRESENT-1PL
'We reap paddy.'

ma:ma: goyam kapayi.
uncle rice-plants cut-PRESENT-3SG
'(Maternal) uncle reaps paddy.'

(2) Spoken (in phonological representation):

mamə/api/ma:ma goyaŋ kapənəwa.
I/we/uncle rice-plants cut-PRESENT
'I/we/maternal uncle reap paddy.'

Phonology

Two notable characteristics of the Sinhala phonological inventory that are shared with Dhivehi but that are different from other Indo-Aryan languages are the lack of aspirated consonant series and a series of prenasalized stops that contrast with nasal-stop clusters. The vowel and consonant inventories are shown in Table 1.

Orthography

Sinhala has an orthographic system of its own, also used for writing Sanskrit and Pali in Sri Lanka. Like

most other South Asian systems it is alphasyllabic; that is, consonants imply an inherent vowel <a>, with other following vowels or the lack of one indicated by satellite diacritics. The independent vowel symbols are used for word initial or independent vowels. Thus ප is <pa> and පා, පි, පු, පෙ, and පො, indicate <pa>, <pi>, <pu>, <pe>, and <po>, with (plain) <p> written as ප්. The full inventory, given in Table 2, does include symbols for aspirate consonants and others for writing Sanskrit and Pali as well as loans from those languages.

Morphology

Sinhala nouns inflect for definiteness, number, and case. The basic gender categories are animate and inanimate. Table 3 gives a partial set for Spoken and Literary. Literary Sinhala also distinguishes masculine and feminine within animate.

Spoken Sinhala has six cases, including the vocative: nominative, dative, genitive, instrumental-ablative, and vocative. Literary Sinhala and some dialects of Spoken also have a distinct accusative, though they differ in form.

Demonstratives and pronouns exhibit a four-way distinction: 1st proximal, 2nd proximal, distal, and (discourse) anaphoric. Thus, roughly, *me*: 'this by me', *oya* 'that by you', *ara* 'that over there', and *e*: 'that has been spoken of'.

As stated earlier, Spoken Sinhala verbs lack person-number-gender agreement, while Literary Sinhala has it for all three categories. Both varieties have a number of forms for tense, mode, voice, and aspect, though the inventories differ somewhat. There is also a three-way derivational system with sets including active, causative, and involitive verbs, though some sets are incomplete. Thus, *kaḍənəwa* 'break' (active, transitive), *kaḍəwənəwa* 'cause (someone) to break', and *kəḍənəwa* 'get broken' (intransitive/involitive). The syntactic/semantic reflexes associated with these forms are complex and involve transitivity, causativity, and involitivity and the case of subjects and other grammatical relations, as well as special characteristics of specific verbs. Thus the causative of *kiyənəwa* 'say, tell' is the common verb for 'read', but it is uncommon, though possible, in its causative sense, and its conjunctive participle *kiyəla* is also the quotation marker/complementizer in Spoken Sinhala (see (6) below).

Syntax

The basic word order in Sinhala is subject-object-verb, though other orders are not only possible but common for pragmatic effects such as foregrounding and emphasis. It is a thorough-going left-branching

Table 1 The Sinhala consonant and vowel inventory

| Consonants | | Vowels |
|---------------------|---|-------------|
| Voiceless stops | k c ʈ t p | i i: u u: |
| Voiced stops | g j ɖ d b | |
| Prenasalized stops | ⁿ g (ⁿ j) ⁿ ɖ ⁿ d ^m b | e e: ə o o: |
| Nasals | ŋ ñ n m | |
| Resonants, spirants | y r l w š s h | æ æ: a a: |

Table 2 The Sinhala writing system

| Vowels | | | | | |
|------------|------------------|----|------------------|----|-----------------|
| අ | ආ | ඇ | ඈ | ඉ | ඊ |
| a | a: | æ | æ: | i | i: |
| උ | ඌ | ඍ | ඎ | ඏ | ඐ |
| u | u: | r | r: | ! | l: |
| එ | ඒ | ඓ | ඔ | ඕ | ඖ |
| e | e: | ai | o | o: | au |
| Consonants | | | | | |
| ක | ඛ | ග | ඝ | ඞ | ඟ |
| ka | k ^h a | ga | g ^h a | ŋa | ⁿ ga |
| ච | ඡ | ජ | ඣ | ඤ | ඦ |
| ca | c ^h a | ja | j ^h a | ña | ⁿ ja |
| ට | ඨ | ඪ | ඬ | ඹ | ඬ |
| ta | t ^h a | da | d ^h a | ṇa | ⁿ da |
| ප | ඵ | බ | භ | ම | ඹ |
| pa | p ^h a | ba | b ^h a | ma | ^m ba |
| | ය | ර | ල | ව | |
| | ya | ra | la | va | |
| ශ | ෂ | ස | හ | ළ | |
| śa | ṣa | sa | ha | la | |

Other

The 'class nasal' ° (Si. *binduva*) is listed following the vowels, but usually represents a velar nasal, transcribed <ŋ> or <m>.

A symbol ො has been introduced for <f>, but often Roman <f> is combined with ප <pa>, as in ෆ.

Table 3 Spoken (colloquial) and literary nouns

| | <i>Singular^a</i> | | <i>Plural</i> |
|----------------------------|-----------------------------|-------------------|-------------------------|
| | <i>Definite</i> | <i>Indefinite</i> | |
| Animate (masculine) | | | |
| Nominative/Direct | | | |
| Coll | miniha 'the man' | minihek | minissu |
| Lit | minisa: | minisek | minissu |
| Accusative | | | |
| Coll | minihawə | minihekwa | minissunwa |
| Lit | minisa: | minisaku (-eku) | minisun |
| Dative | | | |
| Coll | minihata | minihekuṭa | minissunṭa |
| Lit | minisa:ṭa | minisakuṭa | minisunṭa |
| Inanimate | | | |
| Direct ^b | | | |
| Coll and Lit | potə 'the book' | potak | pot |
| Dative | | | |
| Coll and Lit | potəṭa | potakəṭa | potwəlaṭa/ potvəlaṭa |

^aSpoken forms are given in phonological representation; literary forms in transliteration.

^bNouns of this type have no separate accusative in either variety, but only a direct case serving both functions.

(i.e., right-headed) language, and verb and noun modifiers, including relative clauses, precede their heads. (The correlative relative construction generally characteristic of Indo-Aryan languages was lost early

in its history.) It has postpositions, and complementizers are clause final. These characteristics are illustrated in (3) through (6).

(3) siri gunəpa:ləṭə potak dunna.
Siri Gunapala-DAT book-INDEF give-PAST
'Siri gave Gunapala a book.'

(4) mama ada koləmbə iⁿdəla
I today Colombo-GEN from
ko:ciyeŋ a:wa.
train-INSTR come-PAST
'I came from Colombo by train today.'

(5) siri gunəpa:ləṭə dunna potə.
Siri Gunapala-DAT give-PAST-REL book
'The book that Siri gave Gunapala.'

(6) siri i:ye a:wa kiyəla gunəpa:lə kiwwa.
Siri yesterday came COMP Gunapala say-PAST
'Gunapala said that Siri came yesterday.'

Sinhala has the conjunctive participle that is a feature of both Indo-Aryan and Dravidian languages, and it is the major way in which sentence conjunction is effected.

(7) siri kə:mə ka:la gedərə giya.
Siri food eat-CONJPART home went
'Siri ate and went home.'

Nonverbal sentences, which are common, may be of numerous types, of which three are illustrated in (8) through (10). Such sentences do not have a

copula, but vowel-ending adjective predicators take an assertion marker, as in (10).

- (8) Nominal-equational:
 me: potə puskolə potak.
this book ola-leaf book-INDEF
 'This book is an ola-leaf manuscript.'
- (9) Adjectival-attributive:
 me: potə hoⁿdayi.
this book good-ASSMKR
 'This book is good.'
- me: potə alut.
this book new
 'This book is new.'
- (10) Nonverbal modal:
 maṭə e: alut navakəṭa:potə o:nə.
I-DAT that new novel-book want/need
 'I want that new novel.'

Sinhala has the dative subject sentences common in South Asia. A nonverbal example was provided in (10). Dative subject verbal sentences commonly involve involitive verbs, as in (11):

- (11) maṭə aliyek penuna.
I-DAT elephant-INDEF see-PAST
 'I saw the elephant (it was visible to me).'

An uncommon feature of Sinhala is that it also has subjects in case forms other than nominative/direct and dative, in fact, in all except the genitive, as in (12)–(14). (12) illustrates the involitive optative verb inflection, indicating possibility of occurrence.

- (12) minihawə gaⁿgəṭə wæte:wi.
man-ACC river-DAT fall-INVOLPT
 'The man might fall into the river.'
 (There are no accusative subject transitive sentences.)
- (13) ehe: po:lisiyeŋ innəwa.
there police-INSTR be (Animate)
 'There are police there.'
- (14) a:nduweŋ e:kəṭə a:da:rə
government-INSTR that-DAT support-PL
denəwa.
give-PRES
 'The government gives support for that.'

Sinhala has an interesting cleft or focused sentence construction that requires a special form of the verb, as in (15). The focused element may be virtually any type of sentence constituent, and it may be postposed as in (15) but need not be. This structure is very common in discourse and is used in most types of question word questions, as in (16). The question marker *də* that appears in (16) is also the way in which ordinary yes/no questions are formed, as in (17):

- (15) i:ye gunəpa:ləṭə salli
yesterday Gunapala-DAT money
 dunne e: miniha.
give-PAST-FOC that man
 'It was that man who gave Gunapala money yesterday.'
- (16) i:ye gunəpa:lə e: minih aṭə
yesterday Gunapala that man-DAT
 dunne mokak də?
give-PAST-EMPH what Q
 'What did Gunapala give that man yesterday?'
- (17) e: miniha i:ye gunəpa:ləṭə
that man yesterday Gunapala-DAT
 salli dunna də?
money give-PAST Q
 'Did that man give Gunapala money yesterday?'

The related Dhivehi has a similar focus construction. It is also found in several Dravidian languages, and it is very likely that it, like other characteristics such as the completely left-branching nature of Sinhala-Dhivehi, is a result of language contact.

Bibliography

- Coates W & De Silva M W S (1960). 'The segmental phonemes of Sinhalese.' *University of Ceylon Review* 18(3–4), 163–175.
- De Silva M W S (1979). *Sinhalese and other island languages in South Asia*. Tübingen: Gunther Narr Verlag.
- Disanayaka J B (1991). *The structure of spoken Sinhala: 1: Sounds and their patterns*. Maharagama, Sri Lanka: National Institute of Education.
- Fairbanks G W, Gair J W & De Silva M W S (1968). *Colloquial Sinhalese (Sinhala)* (Books 1 & 2). Ithaca, NY: Cornell University South Asia Program.
- Gair J W (1970). *Colloquial Sinhalese clause structures*. The Hague: Mouton.
- Gair J W (1996). 'Sinhala writing.' In Daniels P T & Bright W (eds.) *The world's writing systems*. New York and Oxford: Oxford University Press. 408–412.
- Gair J W (1998). *Studies in South Asian linguistics: Sinhala and other South Asian languages*. Oxford: Oxford University Press.
- Gair J W & Karunatilake W S (1974). *Literary Sinhala*. Ithaca, NY: Cornell University South Asia Program.
- Gair J W & Karunatilake W S (1976). *Literary Sinhala inflected forms: a synopsis with a transliteration guide to Sinhala script*. Ithaca, NY: Cornell University South Asia Program.
- Gair J W & Paolillo J C (1997). *Sinhala (Languages of the world/materials 34)*. München: Lincom.
- Gair J W, Karunatilake W S & Paolillo J C (1987). *Readings in colloquial Sinhala*. Ithaca, NY: Cornell University South Asia Program.
- Geiger W (1938). *A grammar of the Sinhalese language*. Colombo: Royal Asiatic Society.
- Godakumbura C E (1955). *Sinhalese literature*. Colombo: Colombo Apothecaries Ltd.

- Gunasekara A M (1891). *A grammar of the Sinhalese language. Adapted for the use of English readers and prescribed for the Civil Service Examination*. Colombo: Government Press. [Reprinted Sri Lanka Sahitya Mandalaya, Colombo: 1962.]
- Karunatilake W S (1992). *An introduction to spoken Sinhala*. Colombo: Gunasena.
- Karunatilake W S (2001). *Historical phonology of Sinhalese: from old Indo-Aryan to the 14th century AD*. Colombo: S. Godage and Brothers.
- Macdougall B G (1979). *Sinhala: basic course*. Washington D.C.: Foreign Service Institute, Department of State.
- Matzel K & Jayawardena-Moser P (2001). *Singhalesisch: Eine Einführung*. Wiesbaden: Harrassowitz.
- Reynolds C H B (ed.) (1970). *An anthology of Sinhalese literature up to 1815*. London: George Allen and Unwin (English translations).
- Reynolds C H B (ed.) (1987). *An anthology of Sinhalese literature of the twentieth century*. Woodchurch, Kent: Paul Norbury/Unesco (English translations).
- Reynolds C H B (1995). *Sinhalese: an introductory course* (2nd edn.). London: School of Oriental and African Studies. [1st edn., 1980.]

Sino-Tibetan Languages

R J LaPolla, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

The Sino-Tibetan (ST) language family includes the Sinitic languages (what for political reasons are known as Chinese ‘dialects’) and the 200 to 300 Tibeto-Burman (TB) languages. Geographically it stretches from Northeast India, Burma, Bangladesh, and northern Thailand in the southeast, throughout the Tibetan plateau to the north, across most of China and up to the Korean border in the northeast, and down to Taiwan and Hainan Island in the southeast. The family has come to be the way it is because of multiple migrations, often into areas where other languages were spoken (LaPolla, 2001). Proto-Sino-Tibetan (PST) would have been spoken in the Yellow River valley at least 6000 years ago. Waves of migration followed: to the southeast, forming the Sinitic languages, and to the west and southwest, forming the TB languages (the speakers of what became the Bodish languages migrated west into Tibet and then south, all the way to the Bay of Bengal, while the speakers of what became the rest of the TB languages followed the river valleys down along the eastern edge of the Tibetan plateau and across into Burma, India, and Nepal). The large spread of Mandarin Chinese to the northwest, southwest and northeast, giving it its large population and geographic spread, happened only in the last few hundred years.

In the past, and to some extent in China still today (e.g., Ma, 2003), this family was also said to include the Tai-Kadai (Zhuang-Dong) and Hmong-Mien (Miao-Yao) languages of southern China and Southeast Asia, but the resemblances found among Sinitic,

Tai-Kadai, and Hmong-Mien are now understood to be a result of contact influence (these peoples originally inhabited southern China). Sino-Tibetan has the second largest number of speakers of any language family in the world, due largely to the over one billion Sinitic speakers; except for Burmese (see Bradley, 1996), most Tibeto-Burman languages have relatively few speakers.

Subgroupings within ST are still controversial, due to differences in criteria for subgrouping, a paucity of reliable data, particularly on morphosyntactic patterns, and the fact that the development and distribution of these languages has been greatly influenced by migration and language contact. Some of the influential proposals for subgrouping within TB are Grierson 1909, Shafer 1955, Benedict 1972, DeLancey 1987, Sun 1988, Dai, Liu & Fu 1989, Bradley 1997, Matisoff 2003, and Thurgood 2003 (see Hale 1982 for comparison of the older proposals). There is now general agreement on the existence of the following groupings (individual languages listed are only representative; see Matisoff 1996 for the many different names used for TB languages and groupings).

- Qiangic (Qiang, Pumi, Muya, Namuyi, Shixing);
- Lolo-Burmese, comprising the Burmish languages (Burmese, Lawngwaw [Maru], Ngo Chang [Achang], Zaiwa, Lachik [Lashi]) and the Loloish languages (further divided into Northern: Nosu [Yi, Yunnan or Sichuan], Nasu, Nisu; Central: Lahu, Lisu, Nusu, Jinuo; and Southern: Hani, Bisu, Phunoi, Mpi);
- Bodish (Tibetan, Dzongkha, Tamang (several varieties), Tshangla, Takpa);
- Kuki-Chin (Lushai, Asho Chin, Tiddim [Chin, Tedim], Anal, Hmar);

- Bodo-Koch (Bodo, Garo, Dimasa, Kachari, Koch, Rabha);
- Konyak (Tangsa [Naga, Tangsa], Chang [Naga, Chang], Konyak [Naga, Konyak], Nocte [Naga, Nocte], Wancho [Naga, Wancho]);
- Tani (Apatani, Mising [Miri], Adi); and
- Karenic (Pwo [Karen, Pwo], Karenni, Sgaw [Karen, S'gaw]).

There is much controversy over the affiliations of many of the languages of Northeast India and whether they all form a group together (see Burling, 1999; Matisoff, 1999), as well as the positions of the Bai language of Yunnan, China, Newari and the Kiranti languages of Nepal, Dulong-Rawang-Anong (Rawang) of Burma and China, the extinct Tangut language of northwest China, and the rGyalrong language of Sichuan, China, among many others. The latter two are most often said to be part of the Qiangic group, and the Kiranti languages are often seen as forming a higher grouping with the Bodish languages, but LaPolla (2003a), with reference to the morphological paradigms, argued that rGyalrong, the Kiranti languages (Bantawa, Athpare [Athapariya], Dumi, Khaling, Camling), Dulong-Rawang-Anong, the Kham languages, and the Western Himalayan languages (Kinnauri, Rongpo, Chaudangsi, Darmiya; also often grouped with Bodish) should be seen as forming a single higher-level grouping. This grouping was given the name 'Rung' because of the similarity (but not identity) of this proposal to an earlier one by Thurgood (1985). The Rung languages most likely split off from an even higher-level grouping with the Qiangic languages, then rGyalrong split off from the group as migrations moved south, then Western Himalayan split off from Kiranti and Rawang, and then these two groups split (Figure 1; see LaPolla, 2003a, for the evidence).

Within Sinitic, it is generally agreed there are at least six major dialect groups, initially distinguished on the basis of the reflexes of the historically voiced initial consonants (Li, 1936–1937): Mandarin (northern and southwestern China), Wu (Jiangsu and Zhejiang), Xiang (Hunan), Gan (Jiangxi), Yue (Guangdong and Guangxi), and Min (Guangdong,

Fujian, Hainan Island, and Taiwan). The Hakka group of dialects (Guangdong, Fujian, Jiangxi, Sichuan, and Taiwan) is seen by some as part of the Gan group and by others as a separate group. Another three groups were proposed by Li (1987): the Jin group (Shanxi and Inner Mongolia), the Hui group (Anhui and Zhejiang), and the Pinghua group (Guangxi), but these groupings are not universally accepted. Norman (1988, 2003), based on a paradigmatic set of lexical and grammatical items, further grouped the dialect groups into the Northern (Mandarin) group, the Central group (some Xiang dialects, Wu, Gan), and the Southern group (Yue, Hakka, and some Xiang dialects). He left out the Min group because he felt that the Min dialects lay "outside the mainstream of Chinese linguistic development" (2003: 81). That is, they cannot be reconciled with the reconstructed Middle Chinese system (seventh century A.D.) to which the other dialect groups can be traced.

Mandarin has the largest geographic spread and population, and can be subdivided into as many as eight subgroups (see Li, 1987; cf. Ho, 2003), based largely on the reflexes of the stopped tone category. Of these, the Southwestern (Sichuan, Yunnan, Guizhou), Central Plains, and Jianghuai (Southeastern) groups are generally recognized.

One variety of Mandarin, *P<ut - onghu*λa, the 'Common Language' of China today, was developed in the early 20th century (and dubbed *Gulóy<ü*, 'National Language,' at that time), taking the phonology of the Beijing dialect but the lexicon and grammar from a more generalized Mandarin and from the vernacular literature of the time. Standardization and spread of the standard through aggressive educational programs continues today.

Min does not have a large spread and population, but because of the complex nature of its historical development (multiple migrations into the area, causing multiple strata, even within a single variety), it can be subdivided into as many as seven subgroups: Southern, Northern, Central, Eastern, Puxian, Shaojiang, and Qiongwen (Li, 1987). For an excellent book-length synchronic and historical overview of Sinitic, see Norman, 1988; for the best detailed analysis of a single dialect, see Chao (1968).

Proto-Sino-Tibetan was monosyllabic, but with a much more complicated syllable structure than most of the modern languages: **(PREF) (PREF) C_i (G) V (:)* (*C_f*) (*s*) (Matisoff, 1991: 490; *C_i* = initial consonant, *G* = glide, *:* = vowel length, *C_f* = final consonant, *s* = suffixal **-s*; parentheses mark items that do not appear in all syllables). The modern languages have moved much more toward bisyllabic or polysyllabic words, although they are often reduced again to

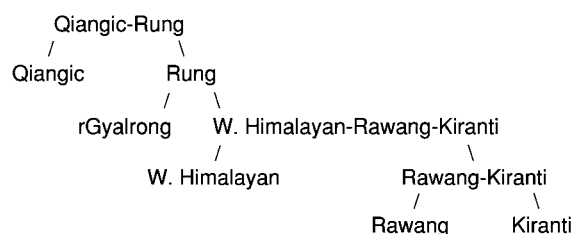


Figure 1 The subgrouping of Qiangic-Rung.

sesquisyllabic (syllable and a half) or monosyllabic forms, and tone systems have developed in Sinitic and many of the TB languages (either through contact, through independent innovation, or a combination of the two). For example, in Sinitic the tones developed out of consonant suffixes (*-s, *-ʔ) and loss of initial voicing (Baxter, 1992: 8.2), and in Lhasa Tibetan the tones developed independently, out of loss of initial voicing and the influence of final consonants. Within this general commonality there is also diversity in phonemic inventories and syllable structures, with, for example, the Qiang language (LaPolla, 2003b) having 36 initial consonants, a complex system of consonant clusters in initial and final position, and no tones, while Lahu (Matisoff, 1973) has only 24 consonant initials, a simple (C)V syllable structure (no consonant clusters), and seven phonemic tones.

Proto-Sino-Tibetan morphology included derivational prefixes and suffixes and a voicing alternation of the initial consonant of some verbs that could affect the valency or form class of a word, but no relational morphology. Many of the modern languages have grammaticalized person-marking affixes on the verb and/or semantic role marking on nouns, but these cannot be reconstructed to the PST level (see LaPolla, 2003a, and references therein). The clause was verb focused, in that the verb was the key element, and noun phrases were optional. This is still the case in most languages. Most have not grammaticalized the kind of constraints on referent identification we associate with the concept of 'subject' and other grammatical relations. If noun phrases appeared in the clause, the verb would have been clause final. In Sinitic the clause is largely verb medial, as the verb has come to function as the divider between topical (preverbal) and nontopical (postverbal) elements (there has clearly been a progressive change away from verb-final order over time). This change has happened to a large extent in Bai and Karen as well. With morphology as with phonology we find diversity of types. Using our examples of Qiang and Lahu again, we find Qiang is agglutinative, whereas Lahu is isolating. Qiang has complex affixal systems of direction marking, person marking, and evidential marking on the verb and definite marking in noun phrases, whereas Lahu has none of these features. Both languages have developed complex sortal classifier systems – a common, but not universal, trait among ST languages. All ST languages have modifier-modified order in noun–noun structures (with genitive-head order being a subtype of this – there was no genitive marking in PST, but some languages have developed genitive marking), as well as relative-head order (Karen has a secondary head-relative order as

well). Proto-Sino-Tibetan had negative-verb order, and this is still true of most ST languages.

Matisoff (2003) grouped the languages in the family into the 'Sinosphere' and the 'Indosphere' due to the linguistic and political influence of China and India, respectively, on the languages. In Indospheric languages, such as the TB languages of Northeast India and Nepal, for example, we often find the development of relative pronouns and correlative structures, and also of retroflex initial consonants. In the Sinosphere we often find the development of tone systems and more analytic structure. We also find contact influence from the Altaic languages in the north (Altaic speakers controlled large parts of northern China for long periods over the last thousand years) and the Austroasiatic, Tai-Kadai, and Hmong-Mien languages in the south. For example, there is a cline from north to south in terms of complexity of tone and also classifier systems (greater in the south, less in the north), and influence on prosody and word structure where the sesquisyllabic light-heavy structure of Austroasiatic languages is also found in many of the southern TB languages, such as Burmese and Jinghpaw (Jingpho), often leading to the reduction of the first syllable in a compound, in contrast to a trochaic stress pattern in northern TB and northern Sinitic, which often leads to the reduction of the second syllable in compounds.

Bibliography

- Baxter W H (1992). *A handbook of Old Chinese phonology*. Berlin & New York: Mouton de Gruyter.
- Benedict P K (1972). *Princeton-Cambridge studies in Chinese linguistics 2: Sino-Tibetan: a conspectus*. Cambridge: Cambridge University Press.
- Bradley D (1996). 'Burmese as a lingua franca (and associated map, #87).' In Wurm S A, Mühlhäusler P & Tryon D T (eds.) *Atlas of languages used for intercultural communication in the Pacific, Asia, and the Americas, vol. II.1*. Berlin: Mouton de Gruyter. 745–747.
- Bradley D (1997). 'Tibeto-Burman languages and classification.' In Bradley D (ed.) *Papers in Southeast Asian linguistics* No. 14, *Pacific Linguistics Series A-86: Tibeto-Burman languages of the Himalayas*. Canberra: Australian National University. 1–71.
- Burling R (1999). 'On Kamarupan.' *Linguistics of the Tibeto-Burman Area* 22(2), 169–171.
- Chao Y R (1968). *A grammar of spoken Chinese*. Berkeley & Los Angeles: University of California Press.
- Dai Q, Liu J & Fu A (1989). 'Guanyu woguo Zangmian yuzu xishu fenlei wenti (On the problem of genetic subgrouping within the Tibeto-Burman languages of China).' *Yunnan Minzu Xueyuan Xuebao* 3, 82–92.
- DeLancey S (1987). 'The Sino-Tibetan languages.' In Comrie C (ed.) *The world's major languages*. New York: Oxford University Press. 799–810.

- Grierson G A (ed.) (1909). *Linguistic survey of India, III, Parts 1–3: Tibeto-Burman family*. Calcutta: Superintendent of Government Printing.
- Hale A (1982). *Research on Tibeto-Burman languages. Trends in linguistics, state of the art report, 14*. Berlin & New York: Mouton.
- Ho D (2003). 'The characteristics of Mandarin dialects.' In Thurgood & LaPolla (eds.). 126–130.
- LaPolla R J (2001). 'The role of migration and language contact in the development of the Sino-Tibetan language family.' In Dixon R M W & Aikhenvald A Y (eds.) *Areal diffusion and genetic inheritance: case studies in language change*. Oxford: Oxford University Press. 225–254.
- LaPolla R J (2003a). 'An overview of Sino-Tibetan morpho-syntax.' In Thurgood & LaPolla (eds.). 22–42.
- LaPolla R J, with Huang C (2003b). *A grammar of Qiang, with annotated texts and glossary*. Berlin: Mouton de Gruyter.
- LaPolla R J & Lowe J B (1994). *STEDT monograph series 1A: Bibliography of the international conferences on Sino-Tibetan languages and linguistics I-XXV*. Berkeley: Center for South and Southeast Asian Languages.
- Li F -K (1936–1937). 'Languages, dialects.' *The Chinese, Year Book*, 121–128. [Reprinted (1973) in *Journal of Chinese Linguistics* 1(1), 1–13.].
- Li R (1987). 'Chinese dialects in China.' In Wurm S A *et al.* (eds.) *Pacific linguistics, series C, no. 102: Language atlas of China parts I and II*. Hong Kong: Longman Group (Far East). [Map and Text A–2.]
- Ma X (ed.) (2003). *Han Zangyu gailun (An introduction to Sino-Tibetan languages)* (2nd edn.). Beijing: Minzu Chubanshe.
- Matisoff J A (1973). *University of California publications in linguistics, 75: The grammar of Labu*. Berkeley & Los Angeles: University of California Press.
- Matisoff J A (1991). 'Sino-Tibetan linguistics: present state and future prospects.' *Annual Review of Anthropology* 20, 469–504.
- Matisoff J A (1996). *STEDT monograph series 2: Languages and dialects of Tibeto-Burman*. Berkeley: Center for South and Southeast Asian Languages.
- Matisoff J A (1999). 'In defense of Kamarupan.' *Linguistics of the Tibeto-Burman Area* 22(2), 173–182.
- Matisoff J A (2003). *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Berkeley, Los Angeles, & London: University of California Press.
- Norman J (1988). *Chinese*. Cambridge: Cambridge University Press.
- Norman J (2003). 'The Chinese dialects: phonology.' In Thurgood & LaPolla (eds.). 72–83.
- Shafer R (1955). 'Classification of the Sino-Tibetan languages.' *Word* 11(1), 94–111.
- Sun H (1988). 'Shilun Zhongguo jingnei Zang-Mian yu de puxi fenlei (A classification of Tibeto-Burman languages in China).' In Eguchi P K *et al.* (eds.) *Languages and history in East Asia: a festschrift for Tatsuo Nishida on the occasion of his 60th birthday*, vol. I. Kyoto: Shokado. 61–73.
- Thurgood G (1985). 'Pronouns, verb agreement systems, and the subgrouping of Tibeto-Burman.' In Thurgood G, Matisoff J A & Bradley D (eds.) *Linguistics of the Sino-Tibetan area: the state of the art*. Canberra: Department of Linguistics, Australian National University. 376–400.
- Thurgood G (2003). 'A subgrouping of the Sino-Tibetan languages: the interaction between language contact, change, and inheritance.' In Thurgood & LaPolla (eds.). 1–21.
- Thurgood G & LaPolla R J (eds.) (2003). *The Sino-Tibetan languages*. London & New York: Routledge.

Relevant Websites

- <http://socrates.berkeley.edu/~ltba/> – *Linguistics of the Tibeto-Burman Area*.
- <http://socrates.berkeley.edu/~jcl2/> – *Journal of Chinese Linguistics*.
- <http://stedt.berkeley.edu/> – Sino-Tibetan Dictionary and Thesaurus Project.
- <http://tibeto-burman.net> – Tibeto-Burman Linguistics Domain.

Siouan Languages

R L Rankin, University of Kansas, Lawrence, KS, USA

© 2006 Elsevier Ltd. All rights reserved.

At the time of earliest contact with Europeans, the Siouan-speaking peoples were found in an arc extending from the northern high plains of North America, east and southward along the prairie-plains border to the mouth of the Arkansas River, with small enclaves farther to the east and southeast. The Siouan

languages generally bear the names of the Native American tribes that speak them.

Subgroups, Locations, and Speaker Statistics

The Siouan languages fall into four major subgroups named after the river valleys where they were spoken in protohistoric times; however, the classification is based on shared linguistic innovations, not geography.

Missouri River Siouan includes Crow, still spoken in southeastern Montana by perhaps 3000 persons of all ages, and Hidatsa in North Dakota, with approximately 60 speakers, all adults. Mississippi Valley Siouan is split into three major groups, Dakotan, Chiwere-Winnebago, and Dhegiha. Dakotan is spoken by over 10 000 persons of all ages in several dialects, including Assiniboiné, Stoney, Teton, Yankton-Yanktonai, and Santee-Sisseton, scattered across northern Nebraska, Minnesota, the Dakotas, Montana, Manitoba, Saskatchewan, and Alberta. The Chiwere dialects are Ioway, Otoe, and Missouria, spoken originally in Iowa, southeastern Nebraska, and northern Missouri. The Missourias took refuge among the Otoes in 1829, and all three tribes were moved to Oklahoma by the 1880s, where today there are perhaps a very few elderly speakers. Winnebago, called Hochunk by its speakers and originally spoken in Wisconsin, is still spoken by adults both there and in Nebraska. The Dhegiha dialects are Omaha-Ponca, Kansa, Osage, and Quapaw. Omaha-Ponca is still spoken by perhaps 50 adults of both tribes in their ancestral home, Nebraska (Omaha), and near Ponca City, Oklahoma (Ponca). Kansa, also called Kaw, originally of northeastern Kansas; Osage, originally of southwestern Missouri; and Quapaw, originally of eastern Arkansas, now in northeastern Oklahoma, no longer have fluent native speakers.

Ohio Valley Siouan is extinct but once comprised several languages, Biloxi in southwestern Alabama, Ofo in Mississippi, and Tutelo-Saponi, Moniton, and Occaneechee in Virginia. There were a few Tutelo speakers living with the Cayuga in Ontario as recently as the early 1980s. The Mandan language, with only one or two speakers in North Dakota, is considered a separate subgroup by the author and a close relation of Mississippi Valley Siouan by some others. The Dakotan, Chiwere, and Dhegiha languages share a certain amount of mutual intelligibility subgroup internally, but there is little or no intelligibility among these subgroups or among other Siouan languages.

External Relationships

The Siouan family is related to the extinct Catawban languages of the Carolina Piedmont. These included Catawba and Woccon and a number of unattested languages said by explorers to have been similar. There is fairly strong recent evidence that Siouan-Catawban is related to Yuchi, originally spoken in Tennessee. Sapir (1929) proposed even more distant links to the Iroquoian and Caddoan language families,

but there is little agreement among specialists on any of these.

Grammatical and Phonological Features

Siouan languages are primarily head-marking, active-stative, subject-object-verb (SOV), i.e., dependent-head languages of moderate morphological complexity. Sapir characterized Dakota Sioux as complex pure relational, with derivational concepts signaled by agglutinating elements and pure-relational (here, pronominal) concepts somewhat fused. Sapir characterized Dakota's overall morphological technique as agglutinative fusional and the degree of synthesis as synthetic to mildly polysynthetic. Siouan languages are among those considered by many linguists to be pronominal argument languages, i.e., the pronominal prefixes on the verb are considered to be the arguments of that verb, not just agreement markers for external arguments. If they are considered agreement markers, then Siouan languages are double-agreement languages, with prefixes for subject and object, or, alternatively, actor, patient/experiencer, along with additional roles such as recipient, locative, and instrumental. Siouan lexical classes include nouns, verbs, pronouns, postpositions, particles, and probably adverbs, but not adjectives. The equivalents of English adjectives are all conjugatable verbs.

Siouan argument structure is of the active-stative type in which the subjects of stative verbs (and some active verbs with experiencer subjects) and objects of active transitive verbs are marked alike, whereas agentive subjects of active verbs (both transitive and intransitive) are marked differently. Siouan languages possess many of the other syntactic orderings that dependent-head languages tend to have (postpositions, main verb–auxiliary verb, possessor–noun (inalienable), and subordinate clause–main clause). All mark person, number, aspect (not tense), mode, and pronominal case in their verb morphologies, and permit noun incorporation. Nominal incorporation is most active in the northern languages: Crow may incorporate entire relative clauses within the verb. Many of the languages have fairly complex phonological inventories, including aspiration, glottalization, and nasalization contrasts for three or four places of articulation among consonants, and length contrasts for five oral and three nasal vowels. Many, if not most, Siouan languages have pitch accent and tend to assign accent to the second mora of words. Phonologists are warned that the practical orthographies, such as those developed by Riggs for Dakota or La Flesche for Omaha, lack detail necessary for phonological analysis.

Future Scholarship

Siouan scholarship is presently flourishing, but much remains to be done. New dictionaries are being or have recently been elaborated for Crow, Hidatsa, Mandan, Dakota, Chiwere, Winnebago, Kansa, Osage, and Quapaw, along with grammars of Crow, Hidatsa, Chiwere, Omaha, Osage, Biloxi, and Ofo. A comparative Siouan dictionary is nearing completion.

Bibliography

- Boas F & Deloria E (1941). *Dakota grammar, vol. XXIII, second memoir. Memoirs of the National Academy of Sciences*. Washington D.C.: National Academy of Sciences.
- Einaudi P F (1976). *A grammar of Biloxi*. New York: Garland Publishing Company.
- Good Tracks J G (1992). *Baxoje-Jiwere-Nyut'aji-Ma'unke: Iowa-Otoe-Missouria language*. Boulder, CO: Center for the Study of the Languages of the Plains and Southwest, Department of Linguistics, University of Colorado.
- Graczyk R (1991). 'Incorporation and cliticization in Crow morphosyntax.' Ph.D. diss., University of Chicago.

- Ingham B (2001). *English-Lakota dictionary*. Richmond, Surrey: Curzon Press.
- Mixco M J (1997). *Mandan*. Muenchen: Lincom Europa.
- Oliverio G R M (1996). 'Grammar and dictionary of Tutelo.' Ph.D. diss., University of Kansas (Lawrence).
- Parks D & Rankin R L (2001). 'The Siouan languages.' In Sturtevant W (ed.) *Handbook of North American Indians*, vol. 13, part 1. Washington, D.C.: Smithsonian Institution Press. 94–114.
- Quintero C (1997). 'Osage phonology and verbal morphology.' Ph.D. diss., University of Massachusetts.
- Rood D S & Taylor A R (1996). 'Sketch of Lakhota, a Siouan language.' In Sturtevant W (ed.) *Handbook of North American Indians*, vol. 17. Washington, D.C.: Smithsonian Institution Press. 440–482.
- Sapir E (1929). 'Central and North American Indian languages.' In *Encyclopedia Britannica*, (14th edn.) vol. 5. Chicago, IL: Encyclopedia Britannica. 128–141.
- Sapir E (1921). *Language. An introduction to the study of speech*. New York: Harcourt, Brace.
- Trechter S (1995). 'The pragmatic functions of gender deixis in Lakhota.' Ph.D. diss., University of Kansas.

Skou Languages

M Donohue, National University of Singapore, Singapore

© 2006 Elsevier Ltd. All rights reserved.

The languages of the Skou family are spoken along the north coast of New Guinea, from the Skou villages east of Humboldt Bay in Indonesia to Barupu west of Aitape in Papua New Guinea. There are 16 known languages in the family, split fairly evenly between three family-level units and one isolate, I'saka (Krisa). Most of the languages are found along the coast, but the orientation of most groups lies inland. Tone and in most cases either unusual consonants or a high number of vowels feature prominently. Tonal contrasts range from three to six on a monosyllabic word, but in all well-investigated cases the domain of tone is the morpheme, not the syllable. Unusual segments found in the family include the palatal lateral dental affricate of Puare (Puari) and the nonback rounded vowels [u] and [ø] of Skou. Contrastive nasalization on either the syllable or the rime is common. Other phonologically marked features include the lack of contrastive nasal consonants in I'saka and the lack of an /s/ phoneme in Skou or many of the Piore River languages.

Morphosyntactically the languages show a lot of variation from one to another, and only some salient features are mentioned here. The basic order is SOV, with postverbal obliques. Case marking is not used, but verbs typically show prefixal agreement for subject, and object agreement, if present, is suffixal. In the western group there is no suffixal agreement, but we do find alternations in the vowel of the verb root that indicate earlier affixation:

ke-fu
3.SING.NF-see.FEM.OBJ
'he saw her'

(where NF stands for nonfeminine) from earlier ke = fu-u. Compare this with Sumo, which has regular suffixation for object:

b-a-chara-u
MOOD-3.SING.MASC-see-3SING.FEM
'he saw her'

Often a language will employ one or more applicatives; typically a goal (beneficiary or direction) or, secondarily, an accompanier has dedicated applicative morphology, whereas instruments are not marked with applicatives but simply appear in the clause.

Many Skou languages show the frequent use of multiple exponence to mark the subject. In Puare we can see double marking for subject on the verb, once by an infix (marked here with angled brackets) and once by a proclitic, as in the sentence:

aro n-s<h>i
firewood 1.DU-chop<1.SING/DU>
'we chopped firewood'

Similarly in Skou we can find verbs with proclitic, prefix, and vowel agreement. The verb /ø/ 'shave' is [təri] when inflected for third-person plural

te-t-lø
3.PL-3.PL-shave<(3).PL>
'they shaved'

Gender is a pervasive feature of the languages. All the languages distinguish at least two genders in the third-person singular pronominal paradigms, and in most cases gender is found elsewhere as well. In Skou, all the dual pronouns, but none of the plural, are differentiated for gender. Barupu (Warupu) and Ramo both distinguish gender in all but the dual pronouns, both free and bound forms. The Serra Hills languages typically mark gender only in the second- and third-person dual pronouns. In Skou itself, a number of nouns obligatorily mark gender. Thus, *ume* 'woman' cannot appear on its own and must take the feminine clitic *pe*, *pe-ume* 'woman', and *āku* 'child' is heard as *pe-āku* 'girl' or *ke-āku* 'boy'.

Bibliography

Cowan H K J (1952). 'Een toon-taal in Nederlandsch Nieuw Guinea.' *Tijdschrift Nieuw Guinea* 13, 55–60.

Donohue M (2002a). 'Negation and grammatical functions in Skou.' In Collins P & Amberber M (eds.) *Proceedings of ALS2002, the 2002 conference of the Australian Linguistic Society*. Sydney: University of New South Wales. Available at <http://www.arts.unsw.edu.au>.

Donohue M (2002b). 'Which sounds change: descent and borrowing in the Skou family.' *Oceanic Linguistics* 41(1), 157–207.

Donohue M (2003a). 'Agreement in the Skou language: a historical account.' *Oceanic Linguistics* 42(2), 479–498.

Donohue M (2003b). 'Morphological templates, headedness, and applicatives in Barupu.' *Oceanic Linguistics* 42(1), 111–143.

Donohue M (2003c). 'The tonal system of Skou, New Guinea.' In Kaji (ed.) *Proceedings of the symposium on cross-linguistic studies of tonal phenomena: historical development, phonetics of tone, and descriptive studies*. Tokyo: Tokyo University of Foreign Studies, Research Institute for Language and Cultures of Asia and Africa. 329–355.

Donohue M & San Roque L (2004). *I'saka*. Canberra, Australia: Pacific Linguistics.

Foley W A (1986). *The Papuan languages of New Guinea*. Cambridge, UK: Cambridge University Press.

Laycock D C (1975). 'Sko, Kwomtari, and Left May (Arai) phyla.' In Wurm S A (ed.) *New Guinea area languages and language study 1: Papuan Languages and the New Guinea linguistic scene*. Canberra, Australia: Pacific Linguistics. 849–858.

Ross M (1980). Some elements of Vanimu, a New Guinea tone language. *Papers in New Guinea Linguistics* 20, 77–109.

Voorhoeve C L (1971). 'Miscellaneous notes on languages in West Irian.' *Papers in New Guinea linguistics* 14, 47–114.

Slavic Languages

L A Janda, University of North Carolina, Chapel Hill, NC, USA

© 2006 Elsevier Ltd. All rights reserved.

The Slavic language group contains three subfamilies: (1) East Slavic, consisting of Russian, Belarusian (Belarusan), and Ukrainian; (2) West Slavic, consisting of Polish, Czech, Slovak, and Sorbian (the latter spoken in Germany and also known as Lusatian); and (3) South Slavic, consisting of Bulgarian, Macedonian, Slovene (Slovenian), and Bosnian/Croatian/Serbian (BCS; formerly known as Serbo-Croatian). The Slavs are believed to have expanded from an

area corresponding to southwestern Belarus/northwestern Ukraine beginning in the 6th century C.E., an event that contributed to the linguistic differentiation of Late Common Slavic (LCS) into the modern Slavic languages. In the late 9th century a Byzantine mission to the present-day eastern Czech Republic yielded translations of liturgical texts into Old Church Slav(on)ic, a written language presumed to be very close to LCS. These documents have made it possible for us to reconstruct the history of the Slavic languages quite reliably. Orthography follows religion in dividing the Slavic languages into an Eastern/Orthodox Christian group that uses the Cyrillic alphabet (Russian, Belarusian, Ukrainian, Bulgarian,

Macedonian, and part of BCS), and a Western/Catholic and Protestant group that uses the Latin alphabet with the addition of diacritics (Polish, Czech, Slovak, Sorbian, Slovene, and part of BCS).

Phonological History

Within the Indo-European language family, the closest relatives to the Slavic languages are the Baltic languages (Latvian and Lithuanian). Both Slavic and Baltic are ‘satem’ languages, a name based on the Avestan word for ‘hundred,’ which identifies the reflex of Proto-Indo-European (PIE) $k' \rightarrow s$ (and $g' \rightarrow z$), as in the Late Common Slavic (LCS) *sъto* ‘hundred’. Peculiar to Slavic (though with some analogues in Baltic and Indo-Iranian languages) is the ‘ruki’ rule sound change, which caused s (\check{s}) $\rightarrow x$ in positions following r , u , k/g , and i , as in Proto-Indo-European (PIE) *ousos* \rightarrow LCS *uxo* ‘ear’. ‘Ruki’ and ‘satem’ are ancient changes in the development of PIE into Early Proto-Slavic (EPSI). The subsequent era linking EPSI and LCS is marked by sound changes that affected all of Slavic, though their ultimate outcomes are not entirely uniform. Many EPSI-to-LCS sound changes reflect a phonotactic strategy aimed at creating ‘ideal’ syllables of rising sonority and level tonality, i.e., syllables with CV structure where both the C and V elements had the same (high or low) tonality (also known as ‘syllabic synharmony’). The conflict between the most normal structure for a root morpheme, which was CVC, and the ideal syllable structure of CV resulted in the great number of morphophonemic alternations so characteristic of Slavic. The last element in a CVC sequence was in a precarious position: either it was assigned to the syllable containing the preceding CV, in which case sonority constraints made it subject to absorption or loss, or it was assigned to the following syllable, where tonality constraints could subject it to mutation. We will look at each group of sound changes separately.

Rising Sonority

Rising sonority motivated syllable shape changes CVC \rightarrow CV and V \rightarrow CV, which resulted in both loss of final consonants, as in EPSI *su:nus* (cf. Gothic *sunus*) \rightarrow *synъ* ‘son’, and prothesis, as in EPSI *esti* (cf. Latin *est*) \rightarrow *jestъ* ‘is’. If a syllable peak contained a diphthong (a vowel followed by a sonorant: a glide, nasal, or liquid), its sonority rose but then dipped, and this lack of conformity to rising sonority also motivated changes in syllable structure, mainly monophthongization or metathesis.

Diphthongs ending in a glide monophthongized to yield new vowels:

- $ei \rightarrow i$, as in *zeim-* \rightarrow *zima* ‘winter’
- $\hat{a}i \rightarrow \check{e}$ (known as ‘jat’), as in *măix-* \rightarrow *měxъ* ‘fur’
- $eu \rightarrow (j)u$, as in *teu-* \rightarrow *tjudъ* ‘alien’
- $\hat{a}u \rightarrow u$, as in *lău-* \rightarrow *luna* ‘moon’.

The subsequent development of ‘jat’ is quite diverse in Slavic.

Diphthongs ending in a nasal monophthongized to yield nasal vowels:

- $e/i + m/n \rightarrow \epsilon$, as in *swent-* \rightarrow *svetъ* ‘holy’
- $\hat{a}/u + m/n \rightarrow q$, as in *zămb-* \rightarrow *zqbъ* ‘tooth’

Polish is the only Slavic language that retains nasality for these vowels (though they have been reorganized: those that developed length became the back nasal q , whereas those that remained short became the front nasal ϵ). The remaining Slavic languages denasalized these vowels, with various results. Thus *svetъ* ‘holy’ and *zqbъ* ‘tooth’ yield, respectively: Russian *sviatoi* and *zub*, Polish *święty* and *zqb*, Czech *svatý* and *zub*, Slovene *svet* and *zob*, BCS *svet* and *zub*, Bulgarian *svet* and *zъb*.

Diphthongs ending in a liquid differed in the presence or absence of an initial consonant and in whether the vowel was a ‘full’ vowel or a reduced vowel (‘jer’), and are referred to as ORT (for orC- and olC-), TORT (for CorC, CerC, ColC, and CelC), and TъRT (for CъrC, CъlC, CъrC, CъlC). Overall, these are referred to as the ‘TORT’ phenomena, and the results (particularly in terms of vowel quality) are quite varied across Slavic. The examples represent only a fraction of the relevant data:

- ORT reflexes show metathesis: *orstъ* ‘growth’ \rightarrow Russian *rost*, Polish *-rost*, Czech *růst*, BCS *rast*, Bulgarian *rast*.
- TORT reflexes show an epenthetic vowel in Russian (pleophony, creating two syllables from one), and metathesis elsewhere: *gordъ* ‘enclosure’ \rightarrow Russian *gorod*, Polish *gród*, Czech *hrad*, BCS *grad*, Bulgarian *grad*.
- TъRT reflexes are the most varied and hard to characterize by rule: *volkъ* ‘wolf’ \rightarrow Russian *volk*, Polish *wilk*, Czech *vlk*, BCS *vuk*, Bulgarian *vъlk*.

Syllabic synharmony

Syllabic synharmony was violated when a low tonality consonant was followed by a high tonality vowel (or sonorant), or when a high tonality consonant was followed by a low tonality consonant. The solution in both cases was to raise the tonality of the low tonality segment. Raising the tonality of consonants yielded the postalveolar fricatives and affricates conspicuous in the Slavic languages, resulting from the palatalizations of velars. In the first palatalization, $k \rightarrow \check{c}$, $g \rightarrow \check{ž}$,

$x \rightarrow \check{s}$ before a front vowel or j uniformly throughout Slavic: *plākĵām* \rightarrow *plačq* ‘I weep’, *gen-* \rightarrow *žena* ‘woman’, *du:xe:tei* \rightarrow *dyšati* ‘breathe’. The second (and third) palatalization of velars took place in two environments: after a high front vowel (or diphthong containing one) or before *ai*. This palatalization yielded $k \rightarrow c$, $g \rightarrow z$ (*dz* in Polish), $x \rightarrow s$ (*š* in West Slavic): *ātikās* \rightarrow *otъcb* ‘father’, *kāinā:* \rightarrow *cěna* ‘price’, *kuningās* \rightarrow *kъnežъ* ‘prince’ (cf. Polish *ksiądz* ‘priest’), *nāgāi* \rightarrow *nozě* ‘leg.DAT/LOC.SG’ (cf. Polish *nodze*), *vixās* \rightarrow *vъsb* ‘all’ (cf. Czech *všechn*), *xāir-* \rightarrow *sěr-* ‘gray’ (cf. Czech *šerý*). The velar palatalizations show a loss (except for Polish) of the stop quality of g , and this was part of a larger phenomenon which included the lenition of g in all positions to a velar or uvular fricative in East Slavic (except Russian), Czech, Slovak, and Upper Sorbian.

Dentals followed by j (and similar clusters) were subject to similar sound changes. Throughout Slavic *sj* \rightarrow *š* and *zj* \rightarrow *ž*: *peisjā:m* \rightarrow *pišq* ‘I write’, *mā:zjā:m* \rightarrow *mažq* ‘I smear’. Original *dj* (also *deu* and *gti*) and *tj* (also *teu* and *kti*), as in LCS *medja* ‘boundary’, *světja* ‘candle’, yielded a variety of reflexes: Russian *mezhal svecha*, Polish *miedzalświeca*, Czech *mezelsvíce*, Slovene *mejalsveča*, BCS *međalsvijeca*, Bulgarian *mezhdalsveshch*. The various palatalizations occur both in roots and at morpheme boundaries, where they occasion various morphophonemic alternations of consonants in Slavic languages. The principle of raising the tonality of a consonant followed by a high tonality vowel has been further continued in some languages: Russian has developed phonemic palatalization, such that all nonpalatal consonants are opposed hard to soft, except the dental affricate *ts*; in Polish this goes one step further and dentals are palatalized to palatals (*t/d/s/z/n* \rightarrow *ć/dź/ś/ź/ń*) before front vowels.

A low tonality vowel following a high tonality consonant (usually j) was also subject to the adjustment of syllabic synharmony, and this resulted in the fronting of back vowels: *mārjās* \rightarrow *morje* ‘sea’, *sjū: tei* \rightarrow *šiti* ‘sew’.

Vowel Distinctions

EPSI had four vowels, all of which could be long or short: *i*, *u*, *e*, *ā*. These vowels were reinterpreted as eight LCS vowels, differentiating the long and short vowels qualitatively. Thus long *i:* \rightarrow *i*, *u:* \rightarrow *y*, *e:* \rightarrow *ě*, *ā:* \rightarrow *a*; and short *i* \rightarrow *ь*, *u* \rightarrow *ъ*, *e* \rightarrow *e*, *ā* \rightarrow *o*.

The LCS era (and the law of rising sonority) comes to an end with the loss of the two short high vowels, *ь* (‘front jer’) and *ъ* (‘back jer’) in weak positions, commonly known as ‘the fall of the jers’. A jer was strong in a syllable preceding a syllable with a weak jer; all

other jers were weak. Weak jers were lost, but strong jers attained the status of full vowels. The fall of the jers created new closed (CVC) syllables, new consonant clusters, and many vowel/zero morphophonemic alternations. In this example, the strong jer is underlined: LCS *sъnъ/sъna* ‘dream.NOM/GEN.SG’ yields Russian *son/sna*, Czech *sen/sna*, BCS *san/sna*.

Prosody

LCS had a system of phonemic pitch and sub-phonemic stress. Although length had been lost in the re-interpretation of vowels, it was subsequently re-established in parts of the Slavic territory. Russian, Belarusian, Ukrainian, and Bulgarian have phonemic stress. BCS and Slovene have phonemic pitch and length. Polish and Macedonian have fixed stress on the penultimate and antepenultimate syllables, respectively. Czech and Slovak have phonemic length and fixed stress on the initial syllable.

Morphological history

Declension

LCS was, and Slavic languages for the most part remain, highly synthetic, with distinct inflectional desinences as well as derivational suffixes and prefixes affixed to roots. EPSI declensions were based mostly on stems with theme vowels, with a few consonantal stems. By the LCS period, the declensions had moved toward association with genders, and the theme vowels were absorbed by sound changes into synthetic desinences that mark case, number, and gender. LCS had three numbers, singular, dual (with restricted case distinctions), and plural, but all the modern languages except Slovene and Sorbian have lost the dual. Slavic maintained much of the PIE case structure, though it merged the ablative with the genitive (restrictive), to yield nominative, genitive, dative, accusative, vocative, locative, and instrumental. The case distinctions (all but the vocative) were subsequently lost in Macedonian and Bulgarian, and the vocative was lost in Russian, Slovene, Slovak, Lower Sorbian, and Belarusian. In addition to the three genders (masculine, feminine, neuter), an animacy distinction developed within masculine during LCS, marked by the substitution of the genitive singular inflection for the accusative singular. Animacy is realized in the plural only by a few languages, in particular Russian and Polish (where it marks only male humans), plus Czech (where it is available only in the nominative plural and dative/locative singular). In LCS, Slavic adjectives were enlarged by the affixation of the corresponding pronominal forms, to create compound adjectives, which

initially signaled definite, as opposed to the shorter 'indefinite' adjectives. BSC and Slovene continue this distribution of long vs. short adjectives. Polish, Czech, and Russian expanded the long compound adjectives, and have restricted the short adjectives to predicate position. Bulgarian and Macedonian maintained only the short adjectives, and developed a postposed article to mark definiteness. The LCS personal pronouns had both long (emphatic) and short (enclitic) forms, and the West Slavic and South Slavic languages continue this distinction.

Conjugation

The most important development for verbs is the evolution of Slavic aspect, which is peculiar because it obligatorily distinguishes perfective from imperfective in all verbal forms, and because the imperfective is more complex and unmarked (whereas it is the marked category in most other languages with this distinction). Aspect is expressed in simplex stems, and via an elaborate system of derivational prefixes and suffixes. The PIE supine, middle, subjunctive, and perfect disappeared in Slavic (but Bulgarian and Macedonian have a new perfect), and the LCS aorist and imperfect tenses have been lost in both East Slavic and West Slavic (except Sorbian). The only two tenses that the modern languages all share are a past (derived from a resultative participle) and a nonpast (usually interpreted as a future if perfective, but as a present if imperfective). Bulgarian and Macedonian lack an infinitive. The Slavic imperative has been innovated from the PIE optative, and the conditional is expressed paraphrastically using an auxiliary from *byti* 'be'. LCS had no distinct future tense, but used instead the perfective nonpast or an auxiliary verb with a participle or infinitive. LCS had a system of four participles expressing present vs. past and active

vs. passive; these survive in their entirety only in Russian. Like its nouns, EPSL verbs were inflected by combining a stem with a theme vowel and a desinence (with the exception of five 'athematic' verbs), and again sound changes obliterated the distinct role of the theme vowel by LCS. In the modern languages, verbs express aspect, tense, person, number, and, in certain forms, gender.

Bibliography

- Carlton T R (1991). *Introduction to the phonological history of the Slavic languages*. Columbus, OH: Slavica.
- Comrie B & Corbett G G (eds.) (1993). *The Slavonic languages*. London/New York: Routledge.
- De Bray R G A (1980). *Guide to the Slavonic languages* (3rd edn.). Columbus, OH: Slavica.
- Gołab Z (1992). *The origins of the Slavs: A linguist's view*. Columbus, OH: Slavica.
- Kondrašov N (1986). *Slavjanskije jazyki*. Moscow: Prosveščenie.
- Meillet A (1965). *Le slave commun*. Paris: Librairie Honoré Champion.
- Schenker A M (1995). *The dawn of Slavic*. New Haven/London: Yale University Press.
- Shevelov G (1965). *A prehistory of Slavic: The historical phonology of Common Slavic*. New York: Columbia University Press.
- Stankiewicz E & Schenker A (eds.) (1980). *The Slavic literary languages: Formation and development* (=Yale East European publications 1). New Haven, CT: Yale Concilium on International and Area Studies.
- Stone G & Worth D (eds.) (1985). *The formation of the Slavonic literary languages* (=UCLA Slavic Studies 11). Columbus, OH: Slavica.
- Townsend C E & Janda L A (1996). *Common and comparative Slavic: Phonology and inflection*. Columbus, OH: Slavica.
- Vaillant A (1950–74). *Grammaire comparée des langues slaves*. Lyon: IAC.

Slovak

R A Rothstein, University of Massachusetts, Amherst, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Slovak is a West Slavic language, most closely related to Czech. It is the native language of some 4.6 million residents of Slovakia, of somewhere between 300 000 and 500 000 residents of the Czech Republic, and of additional speakers in Hungary, Poland, Romania, Serbia, and North and South America.

Orthography

Like other Slavic languages that were historically in the cultural sphere of the Western Church, Slovak uses the Latin alphabet with diacritic marks. Long vowels are marked by an acute accent, *ô* represents the diphthong [uo], and *ä* traditionally represents the vowel [æ], which almost all speakers replace with [ɛ]. The letters *i* and *y* both represent [i], and *í* and *ý* both represent [i:]; the distinction is etymological. Digraphs are used to spell the voiceless velar fricative (*ch*)

and the voiced dental and alveolar affricates (*dz* and *dž*, respectively). The voiceless alveolar affricate and the voiced and voiceless alveolar fricatives are represented, respectively, by *č*, *ž*, and *š*. Palatal stops and sonorants are not specially marked before the vowels *e*, *i*, and *í* or the diphthongs *ia*, *ie*, *iu*; elsewhere they are indicated by diacritics: *t'*, *d'*, *l'*, *ň*. Certain words and categories of words constitute exceptions to this rule for spelling palatals; thus, there is a contrast between the adverb *stále* 'constantly,' pronounced [stál'e], which follows the rule, and the adjective form *stále* 'constant,' pronounced [stále], which does not.

Phonology

The Slovak phonemic inventory consists of fifteen vocalic segments (six short vowels, five long vowels, and four diphthongs) and 27 consonantal ones. Most speakers have only five distinct short vowels, the basic phonetic realizations of which are [i], [ɛ], [ɐ], [ɔ] and [u] (orthographic *i*, *e/ä*, *a*, *o*, *u*), but the different alternation patterns of orthographic *ä* and *e* (the former alternates with *ia*; the latter with *á* or *ia*) are grounds for treating them as representing distinct phonemes. The five long vowels correspond to the short vowels, but /ó/ occurs only in nonnative words, and the distribution of /ě/ outside of nonnative words is limited. The four diphthongs are /uo/, /iel/, /ia/, and /iu/, but the last of these seems always to result from a process of contraction and may therefore not be a phoneme. The diphthongs behave like long vowels with respect to phonologically or morphologically conditioned processes of shortening and lengthening, but the relationships between the long vowels and diphthongs on the one hand and the short vowels on the other is mediated by various contextual constraints. The liquids /r/ and /l/ can be syllabic and in that role distinguish length.

A so-called 'rhythmic law' that operates in Central Slovak dialects and in the standard language mandates the shortening of a syllable that follows a long syllable (one containing a long vowel or a diphthong). Thus, in the masculine nominative singular of adjectives, we find *hladký* 'smooth,' but *krátky* 'short' and *riedky* 'rare.' The rhythmic law does not apply in certain grammatical and derivational contexts, e.g., the declension of adjectives derived from animal names (e.g., *vtáčí* 'bird's').

The consonants that arose from the historical palatalization of velars or from the deiotation of clusters consisting of dental stop or fricative plus glide have lost their palatal character. The historical palatalization of dental consonants, on the other hand, has

given rise to a series of palatal consonants: *t'*, *d'*, *ň*, *l'*. As in most other Slavic languages, final voiced obstruents lose voicing before pause. In obstruent clusters both within phonological words and between words, there is regressive assimilation with respect to voicing. Before a word-initial vowel or sonorant, a word-final obstruent is voiced; such voicing may also occur at morpheme boundaries within words. The laryngeal fricative represented by *h* devoices to the velar fricative *ch*, but when the latter becomes voiced, there is free variation between a voiced velar fricative [ɣ] and *h*. The voiced labiodental fricative *v* behaves like an obstruent at the beginning of a phonological word, but does not cause voicing of a preceding voiceless obstruent within a word. It is realized as [w] in word-final position and is generally realized as [w] word-internally in the environment V_C.

The primary word stress is on the initial syllable and can thus fall on a monosyllabic preposition, especially if the following noun or pronoun is monosyllabic, e.g., *pOd ňou* 'under it,' but also potentially *dO prácy* 'to work.' Unstressed vowels are not reduced, and words longer than three syllables alternate unstressed and secondarily stressed syllables.

Morphology

Nouns distinguish six cases (nominative, accusative, genitive, dative, instrumental, locative); a few masculine nouns have vestigial vocative forms (e.g., *synku* 'son'). Three genders (masculine, feminine, neuter) are distinguished in the singular by agreement phenomena, and a masculine animate subgender can also be distinguished by its syncretism of accusative and genitive and by the ending *-ovi* for dative and locative singular. Certain classes of semantically inanimate masculine nouns also show the accusative-genitive syncretism (e.g., names of trees, mushrooms, diseases: *st'at' duba* 'cut down an oak tree'; *nájsť hriba* 'find a boletus mushroom'; *mat' vreda* 'have an ulcer').

In the plural there is a binary distinction of masculine-personal (nouns referring to male human beings) and nonmasculine-personal (all other nouns); they are distinguished by the nominative endings, by agreement phenomena, and by the accusative-genitive syncretism of the former vs. the accusative-nominative syncretism of the latter. Adjectives and third-person pronouns also distinguish three genders in the singular and two in the plural; the past-tense forms of verbs show a three-way distinction in the singular but have only a single form for the plural. Some nouns have only plural forms (e.g., *vidly* 'pitchfork[s]'); others are used primarily in the singular

(e.g., mass and abstract nouns) but have potential plural forms that usually acquire specialized meanings (e.g., *pivo* 'beer' vs. *pivá* 'kinds or portions of beer'; *láska* 'love' vs. *lásky* 'objects of affection').

Noun declensions are largely gender-based: the masculine and neuter declensions have most endings in common in the singular, while the three feminine declensions in the singular (one for nouns ending in *-a* in the nominative singular and two for those ending in a consonant, i.e., with zero-ending) also share most endings. There is a class of masculine nouns ending in *-a* that refer to male human beings; they follow the masculine declension except for the genitive and accusative singular, which use the *u*-ending of the feminine *a*-declension. In the plural, feminine and neuter nouns have common oblique-case endings, which are different from those of masculine nouns. The traditional presentation of multiple declensional types is based on the fact that certain case endings are dependent on the nature of the final stem consonant – whether it is 'soft' (palatal or 'historically soft,' i.e., the result of historical palatalization or deiotation) or not. Cf. dative singular *žene* vs. *ulici*, which traditional grammars describe as belonging to different declensional types (*žena* 'wife' vs. *ulica* 'street').

Most inherited consonant mutations have been lost from noun declensions. The remaining mutations affect velars and dentals in the masculine personal nominative plural (e.g., *vojak/vojaci* 'soldier[s]'; *Američan/Američania* 'American[s]'), with the alternation /n/ to /ň/; *pilot/piloti* 'pilot[s]'), with /t/ to /t'/ and dentals in the locative singular of all three genders (e.g., *sused/susede* 'neighbor,' *žena/žene, mesto/meste* 'city'). Noun declensions do show quantitative alternations of vowels, for example, between forms with an ending and forms with a zero ending (e.g., NOM/ACC sg *chlieb* vs. GEN sg. *chleba* 'bread,' NOM sg. *ruka* vs. GEN pl. *rúk* 'hand; arm').

Slovak verbs belong to one of two aspectual categories, perfective or imperfective; there are also some biaspectual verbs (e.g., *absorbovať* 'absorb,' *pomstiť* 'avenge'). Perfective verbs express accomplishments or transitions; imperfective verbs express states or activities/processes. Imperfective verbs are typically unprefixated; adding a prefix perfectivizes the verb, sometimes also adding an additional semantic component (e.g., *písať* 'write = engage in the activity of writing'/'*napísať* 'write = get something written' vs. *prepísať* 'rewrite,' *opísať* 'describe,' *popísať* 'write a lot'). There are also productive ways of imperfectivizing a perfective verb through a change in suffix and/or the stem (e.g., *prepisovať* 'engage in the activity of rewriting,' *opisovať* 'engage in the activity of

describing'). Occasionally, corresponding verbs are based on different stems (e.g., imperfective *brať* vs. perfective *vziať* 'take'), and some verbs have no corresponding verb of the opposite aspect (e.g., imperfective *mať* 'have' or perfective *vydržať* 'bear, stand').

Imperfective verbs have synthetic forms for past and present tense and analytic forms for the future tense; perfective verbs form their past tense in the same way as imperfective verbs, but the forms that look like the present-tense forms of imperfective verbs normally express future tense (or, under certain circumstances, potentiality). An analytic pluperfect tense is formed mostly from perfective verbs. The perfective/imperfective distinction is also present in infinitives, imperatives, and conditional/subjunctive forms. The last of these is formed analytically and distinguishes present vs. past ('would X' vs. 'would have X'). Imperfective verbs form verbal adjectives and adverbs expressing simultaneity, while perfective verbs form verbal adjectives and adverbs that express temporal precedence or subordination to the action of the main verb. Both perfective and imperfective transitive verbs form passive participles, which can be used with *byť* 'be' to form passive constructions.

Within the imperfective aspect, a further distinction is made between determinate and indeterminate verbs of motion. The former designate motion in a single direction on a single occasion, while the latter do not have those restrictions and can therefore designate repeated motion, the ability to move, etc. (e.g., determinate *ísť* vs. indeterminate *chodiť*). Many imperfective verbs also have derived iteratives that express repeated, often regular, actions (e.g., *hrávať* 'play frequently' from *hrať* 'play').

Most of the inherited consonant alternations have been eliminated from present-tense paradigms, except for the alternation between palatals and dentals (*idiem* 'I'm going,' *idieš* 'you're going' vs. *idú* 'they're going'; cf. *pečiem, pečieš, pečú* 'I'm baking, etc.,' and the corresponding Polish forms *piekę, pieczesz, pieką*). Other inherited consonant alternations are reflected in the relation between infinitive and present tense (*písať* 'to write' vs. *píšem* 'I write'), past tense and present tense (*mohol* 'he could' vs. *môže* 'he can'), or perfective and derived imperfective (*potvrdiť* 'confirm – pf.' vs. *potvrďovať* 'confirm – impf.'). Quantitative alternations of vowels appear in conjugation (e.g., *piecť* 'to bake' vs. *pečiem*) and in the derivation of imperfective from perfective verbs (e.g., *kúpiť* vs. *kupovať* 'buy' or *skrýť* vs. *skrývať* 'hide'), as well as in nonverbal derivation (e.g., *Nitran* 'man from Nitra' vs. *Nitrianka* 'woman from Nitra').

Syntax

Slovak word order is relatively free and is used, together with sentence intonation, to express the informational structure of the utterance. Thus, the rheme normally follows the theme in emotionally neutral speech. Pronominal and some verbal clitics follow the first stressed word in a sentence. Among them is the particle *sa*, which is historically the enclitic accusative form of the reflexive and reciprocal pronoun. The reciprocal function is still present (e.g., *poznáme sa* 'we know one another'), but true reflexive uses are rare (e.g., *brániť sa* 'defend oneself'). Verbs with *sa* can express a variety of meanings, among other things, a kind of middle voice (e.g., *umývať sa* 'wash/wash up/get washed') and also an intransitive verb with an unaccusative subject (e.g., *lekcia sa začína* 'class is beginning'). They can also be used in passive constructions with unexpressed agent (e.g., *reči sa hovoria a chlieb sa je* 'speeches are spoken, but bread is eaten'). As in some other Slavic languages, *sa* has acquired the function of a generic human subject, parallel to German *man* or French *on*, with third-person singular agreement (e.g., *hovoria sa* 'they/people say').

The enclitic dative form of the reflexive and reciprocal pronoun, *si*, occurs both in its literal meaning (e.g., *pomáhať si* 'help one another' or 'help oneself') and like *sa*, as a component of *reflexiva tantum* (e.g., *všimáť si* 'notice'; cf. *báť sa* 'be afraid'). Both *sa* and *si* also combine with prefixes to produce a variety of Aktionsart meanings (e.g., *nasedieť sa* 'have one's fill of sitting,' *pospať si* 'have oneself a nap').

First- and second-person subject pronouns are normally used only for contrast or emphasis; third-person subject pronouns are typically dropped after their first use, unless a previous theme has been reintroduced. Nonfamiliar address uses second-person plural forms of pronouns and verbs.

Lexicon

In addition to preserving its Common Slavic patrimony, the Slovak lexicon has been open to borrowings and adaptations from neighboring languages. Among the earliest borrowings were elements of Christian terminology from Latin, often via German. Throughout the centuries, those two languages, as well as Czech, Hungarian, and Romanian have been important linguistic donors (Slovak has also contributed to Hungarian); in more recent times, Polish, Russian, French, and English have also served as source languages. In the last decades, the role of internationalisms and Anglicisms has been especially important.

Dialectology

The three major dialect areas are Western Slovak, which is transitional to Moravian Czech; Central Slovak, which served as the basis for the literary language; and Eastern Slovak, which is transitional to Polish. The most striking feature of Eastern Slovak is the lack of quantitative distinctions and a tendency to penultimate stress. Central and Western Slovak both distinguish long and short vowels, but the 'rhythmic law' that (with a variety of systematic exceptions) prevents two successive long syllables applies only in Central Slovak. Central and Western Slovak both have initial stress.

History

The incorporation of the Slovak lands into the Hungarian kingdom that was established at the end of the 10th century separated the Slovaks from the Czechs, who maintained their independence until being subdued by the Habsburgs in 1648. The written language of the Hungarian kingdom, and therefore also of Slovakia, was Latin, but thanks to continuing contacts with their Czech brethren, Slovaks began to use Czech as a written language as well. This was especially true after the establishment of Charles University in Prague in 1348, where Slovaks were among the students, and with the influence of the Hussite movement in the 15th century. From the beginning, the Czech written by Slovaks showed the influence of spoken Slovak.

As early as the 15th century there were efforts to write in Slovak, but the first comprehensive effort to create a Slovak literary standard was by a Catholic priest, Anton Bernolák, at the end of the 18th century, who based his norms on the Western Slovak dialects. Perhaps because Western Slovak is closest to Czech, Bernolák's project did not win general acceptance. Slovak Protestants continued to base their writing on the language of the Czech Kralická Bible.

In the middle of the 19th century, Ľudovít Štúr and his colleagues proposed a new literary standard based on the Central Slovak dialects, and this became the basis of the modern Slovak literary language. During the first Czechoslovak Republic (1918–1938), Slovak linguists had to cope with the official doctrine of a single Czechoslovak language, and after World War II the question of the relationship between the Slovak and Czech languages was still on the agenda. Since the creation of two independent states in 1989, there is evidence of decreasing mutual intelligibility, especially among the younger generation of Slovaks and Czechs, who are less exposed to mass media in the other language.

Bibliography

Kačala J, Eichler E & Šikra J (eds.) (1992). *A reader in Slovak linguistics: studies in semantics*. Munich: O. Sagner.

Rubach Jerzy (1993). *The lexical phonology of Slovak*. Oxford: Clarendon Press.

Short David (1993). 'Slovak.' In Comrie B & Corbett G G (eds.) *The Slavonic languages*. London and New York: Routledge. 533–592.

Slovene

M L Greenberg, University of Kansas,
Lawrence, KS, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Overview

Slovene (or Slovenian), the titular language of the Republic of Slovenia, is spoken by some 2.4 million people, including speakers in bordering areas in Italy, Austria, and Hungary as well as in diaspora communities in Argentina, Australia, Canada, and the USA.

Together with Bosnian, Croatian, and Serbian (Serbo-Croatian), Slovene makes up the Western subgroup of the South Slavic branch of the Slavic languages (Indo-European). Slovene transitions to the Čakavian and Kajkavian dialects of Croatian. It is less close to the Štokavian dialect, the basis for the Bosnian, Croatian, and Serbian standard languages. Ancient connections to the central dialect of Slovak (West Slavic) are evident.

Slovene is traditionally divided into seven dialects, each of which has further dialect differentiation: (I) littoral dialects, spoken partly in Italy; (II) Carinthian, spoken mostly in Austria; (III) Upper Carniolan; (IV) Lower Carniolan; (V) Styrian; (VI) Pannonian, spoken partly in Hungary; (VII) Rovte (Figure 1). There are 48 distinct local varieties.

Standard Slovene is constructed of features from various dialects and historical stages of Slovene and does not correspond exactly to any one dialect. Even in the capital, Ljubljana, everyday speech differs in fundamental ways from the standard; compare standard: *kaj mislite?* 'what do you think?' and colloquial: *kva mislíte?*

Historical Development and Emergence of Literary Language

By the 6th or 7th century A.D., Proto-Slovene was spoken in an area bounded by the Tagliamento River, the Gulf of Trieste, Linz and the outskirts of Vienna, and the southern end of Lake Balaton. The

Proto-Slovene speech territory gradually diminished in the medieval period as speakers shifted to Friulian, Italian, German (Standard German), and Hungarian, leaving a core area today consisting of the Republic of Slovenia plus border areas in Italy, Austria, and Hungary.

The earliest surviving documents are the Freising Folia, liturgical texts composed around 1000 A.D., which are among the oldest attestations of any Slavic language. There are a few surviving Slovene documents dating from then until the middle of the 16th century, mostly religious and legal texts. The first printed books in Slovene are Primus Truber's (1508–1586) *Catechismus* (1550) and Jurij Dalmatin's (1547–1598) translation of the Bible (1584), which mark the first attempt at a standard language. Truber modeled the language on the speech of Ljubljana and his native Lower Carniolan. The Counter-Reformation submerged Truber's legacy, while the Protestants developed a regional literary language in the northeast.

Until the 19th century, Slovene remained secondary to the state language, German (Standard German), and, regionally, Italian and Hungarian. Modern standard Slovene dates to Jernej Kopitar's 1809 grammar, the prestige of which was elevated by the poet France Prešeren (1800–1849) and the intellectual circle around Sigismund Zois (1747–1819). The orthographic system essentially as it is found today was codified in Maks Pleteršnik's Slovene–German Dictionary (1894–95).

Political Issues and Language Maintenance

After the incorporation of the Slovene speech territory (minus border regions in Austria, Italy, and Hungary) into the Kingdom of Serbs, Croats and Slovenes in 1918 (renamed Yugoslavia in 1929), Slovene now became subordinate to Serbo-Croatian, the *de facto* lingua franca of the Yugoslav state. The legal status of Slovene was raised after World War II. Its rights as an official language were reaffirmed in the 1974 Yugoslav Constitution. In reality, the status of Slovene remained unfavorable with

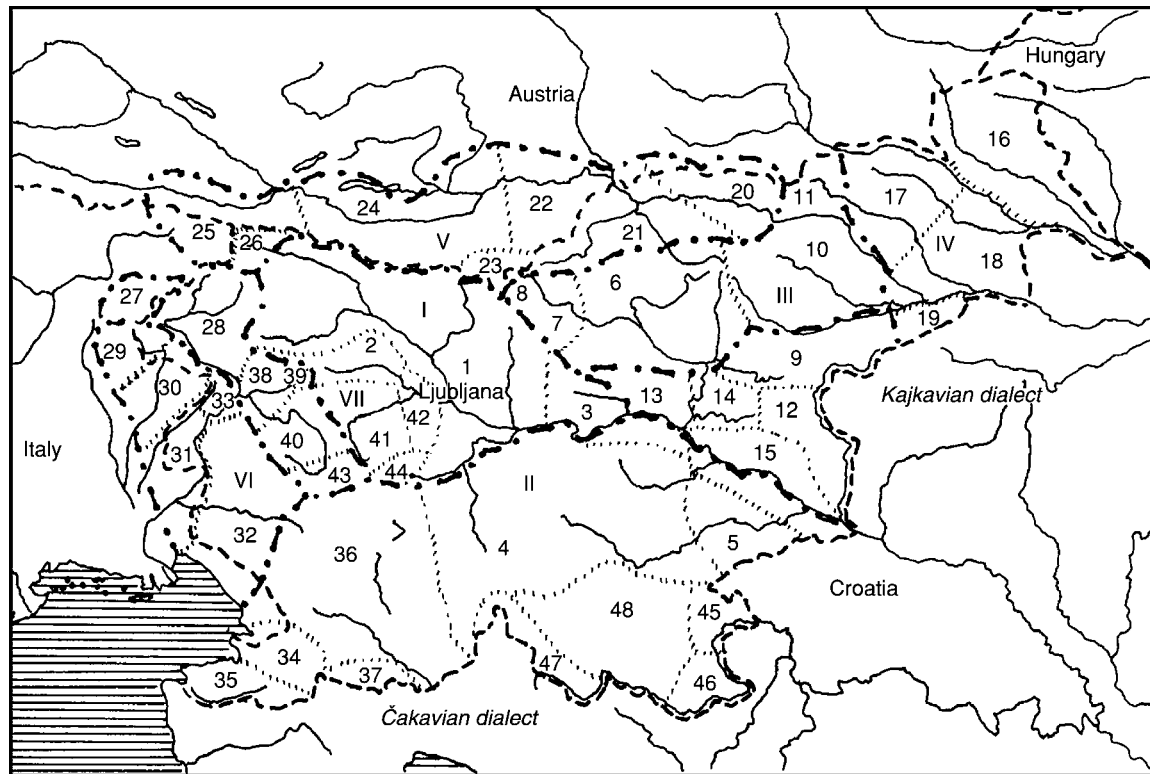


Figure 1 Map of Slovene dialects. Roman numerals indicating local speech varieties are referenced in Greenberg (2000).

respect to Serbo-Croatian, an issue that contributed to Slovene dissatisfaction with Yugoslavia and was resolved with the 1991 Slovene secession from that state.

Slovenes continue to be concerned with language rights among their minorities in Italy, Austria, and Hungary, where they have attempted to encourage the respective governments to accord language rights and allow Slovene-language media and education.

Slovene became one of the official languages of the European Union with the 2004 accession.

Phonology

Writing System

Slovene is written in modified Roman letters, with diacritic marks for sounds not represented by the inherited alphabet (see Table 1). Several other letters are sanctioned in standard orthography to render direct citation of foreign words, viz., Ć, ċ; Ĉ, ĉ; Đ, đ; Q, q; Š, š; X, x; Y, y; Ž, ž; Ž, ž.

Vowel System

See Table 2. *i*, *e*, *ɛ*, *a*, *ɔ*, *o*, *u* occur in long stressed syllables, while stressed *ə* is always short (*pes* [pəs] 'dog'). In unstressed syllables the distinctions

Table 1 The Slovene alphabet

| Upper case | Lower case | Pronunciation (IPA values where significantly different than English) |
|------------|------------|---|
| A | a | [a] |
| B | b | |
| C | c | [ts] |
| Č | č | [tʃ] |
| D | d | |
| E | e | corresponds to tense and lax e-vowels or schwa (see vowel chart) |
| F | f | |
| G | g | |
| H | h | [x] |
| I | i | [i] |
| J | j | [j] |
| K | k | |
| L | l | see explanation under Consonants |
| M | m | |
| N | n | |
| O | o | corresponds to tense and lax o-vowels (see vowel chart) |
| P | p | |
| R | r | tapped or trilled r |
| S | s | |
| Š | š | [ʃ] |
| T | t | |
| U | u | [u] |
| V | v | [w] before a consonant or in word-final position |
| Z | z | |
| Ž | ž | [ʒ] as the s in pleasure |

Table 2 Standard Slovene vowel phonemes

| | Front | Central | Back |
|------------------|-------|---------|------|
| High | i | | u |
| Tense (high-mid) | e | ə | o |
| Lax (low-mid) | ɛ | | ɔ |
| Low | | a | |

between *e*–*ɛ* and *o*–*ɔ* are neutralized to *ɛ* and *ɔ* respectively: *človek* [tʃlɔːvɛk] ‘person-NOM-sing’, *človeka* [tʃlɔːvɛka] ‘person-GEN-sing’; *potok* [pɔːtɔk] ‘stream-NOM-sing’, *potoka* [pɔːtɔka] ‘stream-GEN-sing.’ The grapheme *r* between consonants represents a sequence of *ɔ* + *r*, e.g., *vrt* [vɔrt] ‘garden,’ *srce* [sɔrce] ‘heart.’

Word Prosody

Standard Slovene pronunciation allows two accentual norms, one with pitch accent (characteristic of the Carniolan dialects), the other by stress and vowel length. In the pitch accent system, any long-stressed syllable – almost always only one per accented word – is characterized by either a low rising tone or a high falling tone. Accented words (i.e., not unstressed particles, prepositions, conjunctions, and some pronouns) that lack a long-stressed vowel are short stressed (phonetically high falling) on the final syllable, for example, *brati* [brá:ti] ‘to read’ (low rising), *brat* [brà:t] ‘to go read’ (high falling), *brat* [brát] ‘brother’ (short), *poskòk* ‘hop’ (short).

Stress patterns are morphophonemic in that each morpheme carries an underlying prosodic marker and the concatenation of morphemes to form words determines realization of the placement and identity of the pitch and quantity. The realization of these concatenation rules is that paradigms are characterized either by fixed or by mobile stress patterns, e.g., fixed: *mesto* [mé:sto] ‘town-NOM/ACC-sing’–*mesta* [mé:sta] ‘town-GEN-sing’–*mestu* [mé:stu] ‘town-DAT-sing’; mobile: *meso* [mesò:] ‘meat-NOM/ACC-sing’–*mesa* [mesà:] ‘meat-GEN-sing’–*mesu* [mé:su] ‘meat-DAT-sing.’

Consonant System

See Table 3. *V* is pronounced as English *v* only when it precedes a vowel; otherwise, it is pronounced similarly to *w*: *cerkve* ‘church-GEN-sing,’ *cerkev* [-kəw] ‘church-NOM-sing’; *vrag* [wrak] ‘devil’; *navkreber* [-wk-] ‘uphill.’ *L* is usually pronounced as *w* in word-final position and before a consonant (except in some morphologically conditioned environments, where it is pronounced as [l]): *vedela* ‘she knew,’ *vedel* [-dew] ‘he knew’; *poznavalec* [-ləc] ‘connoisseur-NOM-sing,’ *poznavalca* [-wca] ‘connoisseur-GEN-sing.’

Table 3 Standard Slovene consonant phonemes

| | | Labial | Dental | Palatal | Velar |
|------------|-----------|--------|--------|---------|-------|
| Stops | voiceless | p | t | | k |
| | voiced | b | d | | g |
| Affricates | voiceless | | c | č | |
| | voiced | | | ž | |
| Fricatives | voiceless | f | | š | x |
| | voiced | | | ž | |
| Nasals | | m | n | | |
| Lateral | | | l | | |
| Trill/tap | | | r | | |
| Glides | | v [w] | | j | |

Morphology

Slovene is an inflecting language. Nouns, pronouns, adjectives agree in case, number, and gender. The cases are nominative, accusative, genitive, dative, locative, and instrumental, the last two occurring obligatorily with prepositions. In addition to plural and singular, Slovene has distinct forms for dual. The genders are feminine, masculine, and neuter.

In the standard language masculine adjectives in the nominative and accusative mark the definite article, e.g., *grd obraz* ‘(an) ugly face,’ *grdi obraz* ‘the ugly face.’ In the colloquial language a definite article has developed from a demonstrative pronoun (in all genders and numbers): *grd* ‘ugly’ (generic or indefinite), *ta grd* ‘the ugly (one).’ An indefinite article, also characteristic of colloquial speech, has developed from the numeral ‘one’ (*eden*), e.g., *ena grda faca* ‘an ugly face/guy.’

The present tense of the verb distinguishes person and number. Pronouns are normally dropped unless the subject is emphasized or reference is switched. Second person plural is used also as an honorific for a single addressee.

Verbs distinguish imperfective and perfective aspect, in general, incomplete vs. completed action. Unprefixed verbs are usually imperfective (*pisati* ‘to write’) or bi-aspectual (*nesti* ‘to carry’). Prefixation creates additional, primarily perfective meanings, such as *podpisati* ‘to sign (e.g., a document),’ *odnesti* ‘to carry something away.’ Imperfectives are derived from these prefixed forms by suffixation and sometimes also vowel gradation, e.g., *podpisovati* ‘to sign repeatedly, to be in the process of signing,’ *odnašati* ‘to carry away repeatedly, to be in the process of carrying away.’

Noun and Adjective Inflection

See Tables 4–6.

Table 4 Singular ADJ + noun

| Case | Feminine | Masculine | Neuter |
|--------------|-----------------------------------|--|-----------------------------------|
| Nominative | lepa hiša 'beautiful house' | lep(i) hrib 'beautiful hill, mountain' | lepo mesto 'beautiful town' |
| Accusative | lepo hišo | lep(i) hrib | lepo mesto |
| Genitive | lepe hiše | lepega hriba | lepega mesta |
| Dative | lepi hiši | lepemu hribu | lepemu mestu |
| Locative | (pri) lepi hiši | (pri) lepem hribu | (pri) lepem mestu |
| Instrumental | (z) lepo hišo | (z) lepim hribom | (z) lepim mestom |

Table 5 Plural ADJ + noun

| Case | Feminine | Masculine | Neuter |
|--------------|----------------------|-----------------------|-----------------------|
| Nominative | lepe hiše | lepi hribi | lepa mesta |
| Accusative | lepe hiše | lepe hribe | lepa mesta |
| Genitive | lepih hiš | lepih hribov | lepih mest |
| Dative | lepim hišam | lepim hribom | lepim mestom |
| Locative | (pri) lepih hišah | (pri) lepih hribih | (pri) lepih mestih |
| Instrumental | (z) lepimi hišami | (z) lepimi hribi | (z) lepimi mesti |

Table 6 Dua ADJ + noun

| Case | Feminine | Masculine | Neuter |
|---------------------------|----------------------|-----------------------|-----------------------|
| Nominative, Accusative | lepi hiši | lepa hriba | lepi mesti |
| Genitive | lepih hiš | lepih hribov | lepih mest |
| Locative | (pri) lepih hišah | (pri) lepih hribih | (pri) lepih mestih |
| Dative, Instrumental | lepima hišama | lepima hriboma | lepima mest- oma |

Verb Inflection

The present tense declension marks person and number. Past and future tenses are constructed of an auxiliary (*sem* = past, *bom* = FUT), conjugated as in the present tense, plus a past participle marked for gender and number, e.g., *sem delal* 'I worked-MASC-sing,' *bom delala* 'I shall work-FEM-sing.' The conditional is formed with an invariant particle *bi*, e.g., *bi delali* 'we/you all/they would work' (see Table 7).

Syntax

Word Order

Neutral word order (1) is SVO, but the order may be rearranged depending on emphasis, with either the

Table 7 Present-tense inflection

| | Singular | Plural | Dual |
|---|--------------------|---------|---------|
| 1 | vozi-m '(I) drive' | vozi-mo | vozi-va |
| 2 | vozi-š | vozi-te | vozi-ta |
| 3 | vozi | vozi-jo | vozi-ta |

object moving to the beginning (2) or the subject to the end of the sentence (3).

- (1) *Miran je*
Miran-NOM-sing 3-sing-AUX
kupil
bought-MASC-sing bread.ACC
'Miran bought bread'
- (2) *Kruh je kupil*
bread.ACC 3-sing-AUX bought-MASC-sing
'He bought bread'/'It was bread that he
bought'
- (3) *Kruh je kupil*
bread.ACC 3-sing-AUX bought-MASC-sing
Miran
Miran-NOM-sing
'Miran bought bread'/'It was Miran who bought
bread'

In noun phrases the order is DEM + NUM + ADV + ADJ + noun, where all but the ADV agree in case, number and gender:

- (4) *Tisti dve prav*
those-DEM-NOM-DU-FEM two-NOM-DU-FEM quite-ADV
brihtni punčki
bright-NOM-DU-FEM girls-NOM-DU-FEM
'these two quite bright girls ...'

Clitics

Clitic elements, in accord with Wackernagel's Law, follow directly after the first accented word or noun phrase in the main clause:

- (5) *Trudili smo*
try-IMPERF-PP-MASC-PL AUX-1-PL
se jo razumeti
REFL-PART PRO-3-sing-ACC-FEM understand-INF
'we were trying to understand her'

Subordinate clauses are typically introduced by *da* 'that,' *ki / kateri* 'which,' *ker* 'because,' *ko(t)* 'as,' *če* 'if':

- (6) *Prepričana sem, da je*
convinced-FEM-sing be-1-sing that be-3-sing
tvoj računalnik zastarel
your-MASC- comp-MASC- superannuated-
sing-nom sing-NOM MASC-sing-NOM
'I'm convinced that your computer is obsolete'

- (7) *Pazi,* *ker* *te*
 watch out-IMP-2-sing because PRO-2-sing-ACC
bo *avto* *povozil*
 FUT-AUX-3-SG car-ACC-SG run over-PP-MASC-SG
 'watch out or the car will run you over'

Lexicon

Historical influences on Slovene have come from Friulian, German (Standard German) (especially the Bavarian and Tyrolean dialects), Hungarian and Croatian (Serbo-Croatian), as well as Venetian Italian (Venetian), Dalmatian and Istrian Romance. A number of languages, including Illyrian and continental Celtic, may have made up substrata to Proto-Slovene (or, more likely, to the Romance dialects that preceded it) and are recognizable as trace elements in the vocabulary, e.g., from Celtic *Karavanke* 'Karawanken Alps,' *Kranj(ska)* 'Carniola.' German (Standard German) and English are the source of most contemporary loans, though these are officially deprecated in favor of native formations, which are increasingly accepted in everyday speech, e.g., *zgoščenska* 'compact disk' from *z gostiti* 'to make compact,' replacing *cedejka*. The youngest generation uses English freely, e.g., *ful dober* 'really good' (from Eng. *full*).

Bibliography

- Greenberg M L (ed.) (1997). *The sociolinguistics of Slovene*. International Journal of the Sociology of Language 124.
 Greenberg M L (2000). *A historical phonology of the Slovene language*. Heidelberg: Universitätsverlag Winter.
 Herrity P (2000). *Slovene: a comprehensive grammar*. London & New York: Routledge.
 Lencek R L (1982). *The structure and history of the Slovene language*. Columbus, OH: Slavica.
 Orešnik J & Reindl D F (eds.) (2003). *Slovenian from a typological perspective*. Sprachtypologie und Universalienforschung 56/3 .
 Priestly T M S (1993). 'Slovene.' In Comrie B & Corbett G (eds.) *The Slavonic Languages*. London: Routledge. 388–451.
 Rigler J (1986). 'The origins of the Slovene literary language.' In Rigler J & Jakopin F (eds.) *Razprave o slovenskem jeziku*. Ljubljana: Slovenska Matica. 52–64.
 Snoj M (2003). *Slovenski etimološki slovar*. Ljubljana: Modrijan.
 Stankiewicz E (1980). 'Slovenian.' In Schenker A & Stankiewicz E (eds.) *The Slavic Literary Languages*. New Haven: Yale Concilium on International and Area Studies. 85–102.
 Toporišič J (2000). *Slovenska slovnica*. Maribor: Obzorja.

Sogdian

P O Skjærvø, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Sogdian, an Eastern Middle Iranian language, was spoken at least up to the 8th century in Sogdiana, the area of modern Uzbekistan that includes the cities of Samarkand and Bukhara. Many Sogdians were merchants, however, and traveled east as far as China, bringing with them the Sogdian language. The Manicheans and Christians, as they fled from persecutions from the 3rd century on, took the Sogdian language with them to the farthest reaches of Chinese Turkestan and beyond, into Mongolia, where the Sogdian alphabet was adopted by the local Turks and the Mongolians, who still use it. The Sogdian written remains consist of religious and nonreligious texts. Most of the religious texts are translations, the Buddhist texts from Chinese, the Manichean ones from Persian and Parthian, and the Christian ones from Syriac.

We have Sogdian texts in five different alphabets: Old Sogdian Aramaic, Sogdian-Uighur (Uyghur), Manichean, Nestorian Christian, and Northern Brahmi. The Sogdian Aramaic script is used in the *Ancient letters* (see below) and in graffiti on rocks along the Karakorum Highway in northern Pakistan. The Sogdian-Uighur script is the most common, being used for secular documents, as well as for Buddhist and Manichean texts. The Manichean and Nestorian scripts were used for Manichean and Christian texts, respectively. There are a small number of late Sogdian manuscripts from Turfan written in Northern Brahmi script.

In early times, the Sogdians must have been the neighbors of the Tocharians (see **Tocharian**), who borrowed numerous (proto-)Sogdian words. The modern Iranian language Yaghnobi is the descendant of a Sogdian dialect different from the known Sogdian.

The oldest Sogdian texts are the *Ancient letters*, written on paper and discovered by the British-Hungarian discoverer and archeologist Marc Aurel Stein in eastern Chinese Turkestan (now in The

British Library). The letters can be dated to the early 4th century by references to current events.

From the 8th century, we have a collection of letters and administrative, economic, and legal documents written in the Sogdian script from the archives of King Dhewastich found at Mount Mug east of Samarkand.

The largest corpus of Sogdian texts are the Buddhist texts removed from a cave at Dunhuang in eastern Xinjiang by Aurel Stein and the French scholar and archeologist Paul Pelliot (now in The British Library and the Bibliothèque Nationale). Numerous Sogdian Manichean and Christian texts were discovered at Turfan in northeastern Xinjiang by German archeologists (now in the Brandenburgische Akademie der Wissenschaften in Berlin).

Sogdian phonology and morphology are both conservative and innovative. The most important innovation is the 'rhythmic law,' by which words with long vowels before the endings ('heavy' stems), lose final short vowels. Thus, OIran. SING NOM **wrk-ah* and ACC **wrk-am* 'wolf' are Sogd. *wər-k-i* and *wər-k-ū* ('light' stem), while OIran. **daiw-ah* and **daiw-am* and **daiw-am* 'demon' are both *ḏēw*. Sogdian shares with Ossetic the plural suffix *-t* (originally a collective noun, hence declined like a feminine singular), for instance, *ḏēw-t* 'demons'; forms of *ḏβar-* 'door': SING NOM *ḏβar-ī*, LOC *ḏβar-yá*, PLUR NOM-ACC *ḏβar-t-á*, GEN-DAT, LOC *ḏβar-t-yá* 'at the doors'. Sogdian uses demonstrative pronouns as definite articles (*xō mártī* 'the man', *xā strīš-t* 'the women' [*< strīč-*], *uya kánθ-ī* 'in the city' [LOC]).

The verb system is complex. There are three stems: present, past, and perfect (perfect participle = past stem + suffix *-ē*, FEM *-č-a*; e.g., PRES *pətsāč-* 'fit', PAST *pətsayt-*, PERF MASC *pətsayt-ē*, FEM *pətsayč-á* [*-yt-č->-γ-č-*]). It has all the Old Iranian moods (indicative, imperative, subjunctive, optative, injunctive), as well as active and middle. It has, in modified form, the old imperfect, for instance, PRES *βar-ám*, IMPERF *βar-ú* 'I carry, carried', PRES *wén-am*, IMPERF *wén* 'I see, saw', PRES *θaβr-ám*, IMPERF *θāβr-u* 'I give, gave'. Progressive tenses are formed with the suffix *-skun* (*-sk*) and the future with the suffix *-kām* (*-kan*, *-k*) from a noun meaning 'wish' (IMPERF PROG *βar-á-skun* 'he was carrying', FUT *βar-ám-kām* 'I shall carry'; Christian Sogd. PRES PROG *γərb-ám-sk* 'I am seizing', FUT *wāb-t-kan* 'he shall say').

There is a large range of past tense forms built on the remade Old Iranian perfect system: transitive active tenses with past stem plus the verb *ḏār-* 'hold, have' (e.g., *uyt-u-ḏār-t* 'he has said'), but intransitive and passive tenses with past stem plus copula (e.g., *tγat-ēš* 'you entered', *āžit-əsθa* 'you were born').

The perfect is made with the perfect participle in the same way (e.g., *βast-ē ḏār-and* 'bind-PERF.MASC hold.PRES-3RD.PLUR' = 'they hold/keep bound', *βast-č-á astí* 'bind-PERF.FEM COP.3RD SING' = 'she is (now) bound'). The passive is made with the perfect participle plus 'be, become' (e.g., *βast-ē-t uβ-and* 'bind-PERF.MASC.PL become.PRES-3RD.PLUR' = 'they are being bound', *ánxast-ē əkt-ēm* 'goad-PERF.MASC become.PAST-COP.1ST SING' = 'I was goaded').

Among special formations, note the 'potentialis,' formed with a past participle with the ending (light) *-a* and the verbs *kun-* 'to do' (active) and *β-* 'become' (passive), by which possibility and completion of action are expressed (e.g., *nē žayd-á kun-am* 'NEG uphold.PART do.PRES-1ST.SING' = 'I cannot uphold', *nē āpāt βō-t* 'NEG reach.PART become.PRES-3RD.SING' = 'it cannot be reached', *čānō xwart xurt kun-and* 'when food eat.PART do.IMPERF-3RD.PL' = 'when they had eaten').

There are minor dialect differences between texts written in the Sogdian, Manichean, and Nestorian scripts (e.g., Sogd. *wan-*, *kun-* 'to do', Man., Chr. *kun-*). Christian Sogdian also has phonetically more developed forms (see also on the progressive and future above), e.g., **kərtu-ḏār-am* 'I did, I have done' > Buddhist Sogdian *əktu-ḏār-am* > Christian Sogdian *k-θār-am*.

Bibliography

- Gershevitch I (1954). *A grammar of Manichean Sogdian*. Oxford: Blackwell.
- Henning W B (1958). 'Mitteliranisch.' In *Handbuch der Orientalistik I: Der Nahe und der Mittlere Osten IV: Iranistik 1: Linguistik*. Leiden-Cologne: Brill. 20–130.
- Sims-Williams N (1985). *Berliner Turfantexte 12: The Christian Sogdian manuscript C2*. Berlin: Akademie-Verlag.
- Sims-Williams N (1989). 'Sogdian.' In Schmitt R (ed.) *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert. 173–192.

Somali

F Serzisko, University of Cologne, Cologne, Germany

© 2006 Elsevier Ltd. All rights reserved.

Somali is a Cushitic language of the Afro-Asiatic language family spoken by approximately 10 million speakers in and around Somalia. There are five major Somali dialects (Lamberti, 1986).

Phonology

The following consonant phonemes are distinguished as shown in Table 1.

There are five vowels: a, e, i, o, u and vowel length, in standard orthography indicated by doubling the vowels, is distinctive.

The following shows the letters used in Standard Somali orthography

| | | | | | | |
|--------------------|---|---|---|----|----|----|
| IPA | ʔ | ʃ | h | f | q | x |
| Somali Orthography | ' | c | x | sh | dh | kh |

Tone appears to be distinctive in Somali both on the lexical level and on the grammatical level. It is, however, still a matter of debate whether this tonal distinction is really a tonal distinction or pitch accent (Hyman, 1981 for discussion).

In terms of syllable structure, no word-initial or -final consonant clusters are allowed.

Morphology

Verbs

Somali has a rather rich system of lexical affixes by which new stems can be derived. The main derivational affixes are listed below.

| | | | |
|-----------------|-------------|--------------------|------------------------|
| Causative | | | |
| in jabay | 'is broken' | <i>jab-i-yey</i> | 'to break' |
| Stative/passive | | | |
| am jeex | 'to tear' | <i>jeex-an'</i> | 'to be torn' |
| Autobenefactive | | | |
| an wádayaa | 'to drive' | <i>wadá-na-yaa</i> | 'to drive for oneself' |

The following examples demonstrate the use of these derivational affixes with the verb *fur* 'to open':

| | |
|----------------------|--------------------------------------|
| <i>Wuu fúrayaa</i> | 'He is opening it.' |
| <i>Wuu fúrmayaa</i> | 'It's getting opened, it is opening' |
| <i>Wuu furánayaa</i> | 'He is opening it for himself' |

Morphosyntactic categories of the verb are tense, aspect, and person. There is also an inflectional distinction between main and subordinate predications.

The basic tense distinction is past/non-past, whereby non-past usually has a habitual meaning. Future is expressed by periphrastic construction with the auxiliary *doon* 'want'. There is furthermore an aspectual distinction between progressive and non-progressive.

| | Past | Non-Past |
|-------------------|----------------|---------------------------------|
| Non-Progressive | | |
| <i>Keen-ay</i> | 'brought' | <i>Keen-aa</i> 'brings' |
| Progressive | | |
| <i>Keen-ay-ay</i> | 'was bringing' | <i>Keen-ay-aa</i> 'is bringing' |

The progressive form is historically derived from an auxiliary construction with the verb *hay* 'to have'.

The verb agrees with the subject in gender and number. Inflection is mainly done by suffixes (weak verbs) but there is a small group of five verbs that still have at least partly prefix conjugation (strong verbs). The following sample shows the main verbal forms for the weak and the strong verbs. It should be noted that at least in the main paradigms the forms for 1st and 3rd person masculine and the forms for 2nd and 3rd person feminine are identical.

A sample paradigm for the simple past is given for *keen* 'bring' and *yimi* 'come'

| | Weak verbs | Strong verbs |
|---------|-----------------|-------------------|
| | 'bring' | 'come' |
| 1st/3sm | <i>keenay</i> | <i>imid</i> |
| 2nd/3sf | <i>keentay</i> | <i>ti-mid</i> |
| 1pl | <i>keenmay</i> | <i>nimid</i> |
| 2pl | <i>keenteen</i> | <i>timaaddeen</i> |
| 3pl | <i>keeneen</i> | <i>yimaaddeen</i> |

Predicate negation is expressed by preverbal particles and verbal inflection. An invariable form is used for all persons in the past. The present form inflects regularly.

| | | |
|----------|--------------------|------------------------------|
| Past: | <i>Má keenín</i> | 'I/you/he etc. didn't bring' |
| | <i>má iman</i> | 'I/you/he/etc. didn't come' |
| Present: | <i>má keenó</i> | 'I don't bring' |
| | <i>má imaaddó</i> | 'I don't come' |
| | <i>Má keentó</i> | 'she doesn't bring' |
| | <i>má timaaddó</i> | 'she doesn't come' |

Table 1 Consonant phonemes

| | | | | |
|---|-----|---|-----|----|
| b | t d | q | k g | q |
| m | n | J | | |
| | l | | | |
| | r | | | |
| f | s | f | x | hʃ |
| w | | y | (W) | h |

Nouns

Nominal morphosyntactic categories are case, number, and gender. There is a twofold gender distinction based on masculine and feminine. Gender is marked by tonal distinction and by agreement on determiners like possessives, demonstratives, articles, and the verb. The masculine marker is basically *k* the feminine marker *t*. For a discussion of the status of these, see Lecarme, 2002.

Nin-ka 'the man' *naag-ta* 'the woman'
inan-kayga 'my son' *gacan-tayga* 'my arm'

The basic distinction with number is singular/plural. Plural is marked by several means depending on the length and the gender of the noun. The following examples demonstrate the diverse plural forms.

| | Singular | Plural |
|---------------|---------------|-------------------|
| Woman | <i>náag</i> | <i>naago</i> |
| Road | <i>dariiq</i> | <i>dariiqyo</i> |
| Shoulderblade | <i>gárab</i> | <i>garbo</i> |
| Man | <i>nín</i> | <i>niman</i> |
| Bull | <i>díbi</i> | <i>dibí</i> |
| Story | <i>shéeko</i> | <i>sheekóoyin</i> |
| Father | <i>áabbe</i> | <i>aabbayaal</i> |

There are mainly two cases: subject case and absolutive. The absolutive is the base form and the subject case is marked by a tonal distinction and optionally by the segmental elements *-i* with indefinite and *ku/tu* with definite nouns. This subject marking occurs at the end of the entire noun phrase:

Buug-ga *cusub* *ee* *wiil-kan-i*
 Book-DET new COORD boy-DEM-SUBJ
wux-uu *yaala* *miis-ka* *guud-kiisa*
 wax-3sm located table-DET top-poss3sm
 'The new book of this boy is lying on the table'

Possession is marked on the noun by pronominal suffixes: *aabe-hiis-a* 'his father'. If the possessor is expressed by a noun then the two nouns are juxtaposed and the order is possessee possessor: *faras-ka nin-ka* (horse man) 'the horse of the man'. Alienability is not expressed in Somali, with the exception that kin relationships cannot be simply juxtaposed but use an inverted construction where the possessor is additionally expressed: *ínanka aabi-hiis* 'the father of the boy' (lit. the boy his father). This construction is optional with other possessive relations.

Numerals are nouns and within the noun phrase they precede the noun and actually function as the head of the complex construction.

Laba nín *sáddex nín* *áfar nín*
 'two men' 'three men' 'four men'

Adjectives

Qualitative concepts are expressed by elements whose status is not entirely clear (For a discussion, see the contributions in Bechhaus-Gerst and Serzisko (eds.), 1988). In predicative usage these elements occur with the copula:

Buug-gan waa wanaagsán yahay.
 Book-DEM DM good COP:3sm
 'This book is good.'

In attributive usage, adjectives may agree with their head noun in number; agreement is indicated by reduplication of the first syllable:

Guri cusub 'a new house'
guriyo cuscusub 'new houses'

The comparative is expressed by means of verbal case particles:

Nín-kanu nín-káas wuu ká wèyn yahay
 Man-this man-that DM than(ABL) big COP
 That man is bigger than that man.

The superlative is formed with the preverbal particle cluster *ugú*:

Nínkanu wuu ugú dhèer yahay
 Man-this DM most tall COP
 'This man is the tallest'

Syntax

The structure of a simple sentence can roughly be described as consisting of a verb complex, which contains all the necessary information, and noun phrases, which stand in a kind of appositive relation to this verbal complex. The structure of the verbal complex is as follows:

waa Impersonal object pronoun case marker
 directional verb stem

waa is a declarative marker that stands in complementary distribution with the negative marker. The declarative marker in main clauses, which has also been described as a verbal focus marker (see below), is the left-most element in the verbal complex. The next position may be filled by an impersonal marker. Object pronouns for 1st and 2nd person follow. The 3rd person object is always zero. These pronouns combine with the following case markers. Four cases are distinguished: benefactive /u/, locative/instrumental /ku/, ablative /ka/ and comitative /la/. The following shows the combinations of object pronouns and case marker for the singular pronouns:

| | Benefactive | Loc/Instr | Ablative | Comitative |
|-----------|-------------|-------------|------------|-------------|
| 1st sing. | <i>ii</i> | <i>igu</i> | <i>iga</i> | <i>ila</i> |
| 2nd sing. | <i>kuu</i> | <i>kugu</i> | <i>kaa</i> | <i>kula</i> |

Directional particles indicate whether the action is directed toward the speaker /soo/ or away from speaker /sii/, as in the following examples.

| | | | |
|--------------|---------------|----------------|-----------------------|
| <i>w-aan</i> | <i>ku</i> | <i>ark-ay</i> | 'I saw you' |
| DM-1sg | 2sgOBJ | see-1sg | |
| <i>w-aan</i> | <i>ku-gu</i> | <i>ark-ay</i> | 'I saw you in it' |
| DM-1sg | 2sgOBJ-LOC | see-1sg | |
| <i>w-uu</i> | <i>ku soo</i> | <i>noqd-ay</i> | 'He came back to it.' |
| DM-3sm | LOC back | came-3sg | |

The negation particle also occurs within the verbal complex:

| | | |
|------------------------------|------------|----------------|
| <i>I-i-ma</i> | <i>soo</i> | <i>iibinin</i> |
| 1sgOBJ-DAT-NEG | DIR | bought-NEG |
| 'He didn't buy them for me.' | | |

Sentence Structure

The most striking feature of Somali is the use of focus particle. Noun focus is expressed by the particles *ayaa/baa*, which are alternants whose use is determined by regional and stylistic factors, following the noun in focus. The particle *waa*, which has been described as a declarative marker above, is by some authors called verbal focus marker. Nominal and verbal focus markers stand in complementary distribution, i.e., there can only be one focus marker in a main clause. The form of the focus marker is as a rule dependent on whether the noun in focus is the subject of the sentence or not. If the subject is in focus the marker occurs in its simple form and the predicate occurs in the restricted form. This indicates that the source of the focus construction may be a relative clause. If a non-subject is focused the subject pronoun combines with the focus marker, which yields the following paradigm of forms:

| | Singular | Plural |
|-----------|---------------|------------------------|
| 1st sing | <i>ay-aan</i> | <i>ay-aynu/ay-aanu</i> |
| 2nd | <i>ay-aad</i> | <i>ay-aad</i> |
| 3rd masc. | <i>ay-uu</i> | <i>ay-ay</i> |
| 3rd fem | <i>ay-ay</i> | |

There is, furthermore, a presentative marker *waxa*, which also attracts the subject pronoun. This construction is used to highlight a nominal participant in a kind of clefting construction. The highlighted noun phrase occurs after the verbal complex.

| | |
|-------------------------------|----------------------|
| <i>Shaah b-aan doonayaa.</i> | 'I want some tea' |
| <i>waxaan doonayaa shaah.</i> | 'What I want is tea' |

Word Order

The unmarked word order in a main clause is SOV, the order of the nominal participants is relatively free and interacts with the focus marking system.

| | |
|-------------------------------------|--------------------------------------|
| <i>Nin-kii libaax-ii ayuu dilay</i> | 'The man killed the lion.' |
| Man-DET lion-DET FOC-3sg kill-3sg | |
| <i>Libaax-ii ninkii ayaa dilay</i> | 'The lion was killed by the man' |
| <i>Nin-kii ayaa libaax-ii dilay</i> | 'It was the man who killed the lion' |

Since there is an obligatory syntactic and morphological marking of aspects of discourse structure, Somali can be considered to be an example of a discourse configurational language (Svolacchia *et al.*, 1995).

Questions

Yes-No questions are formed by replacing the declarative marker *waa* by the question particle *ma*:

| | |
|------------------------|-----------------------|
| <i>Cali wuu yimid.</i> | <i>Cali ma yimid?</i> |
| 'Ali came' | 'Did Ali come?' |

If the sentence contains a nominal focus the question particle is placed before the focused noun phrase:

| | |
|-------------------------|----------------------------|
| <i>Cali baa keenay.</i> | <i>Ma Cali baa keenay.</i> |
| 'ALI brought it.' | 'Did ALI bring it?' |

WH-questions always involve nominal focus and the questioned noun phrase stands with the interrogative article *kee/tee*:

| | |
|-------------------------------|--|
| <i>Ninke ayaa yimi?</i> | 'Which man came?' |
| <i>Xaggee buu tegay?</i> | 'Which place did he go? = Where' |
| <i>Sidee baad u sameysey?</i> | 'In which manner did you do it? = How' |
| <i>Intee baad joogaysaa?</i> | 'What amount did you stay? = How long' |

Complex Sentences

Complex sentences can be coordinated or subordinated. Clauses may be coordinated by the particle *oo* as in:

| |
|---|
| <i>Cali hilibkii ayuu keenay oo wàanu cunay</i> |
| 'Ali brought the meat and we ate it.' |

Or they may be conjoined by attaching an element *-na* to the first element of the second clause:

| | | | | |
|---------------------------------|--------|-----------|---------|---------------|
| Cali | w-uu | I | arkay | w-uu-na |
| C. | FOC-3s | 1sgOBJ | see-3sg | FOC-3sm-COORD |
| i-lá | | had | lay | |
| 1sgOBJ-COM | | speak-3sm | | |
| 'Ali saw me and he spoke to me' | | | | |

In subordinated clauses, there is no classifier or focus particle and the verb occurs in its subordinated form.

There is a formal distinction between restrictive and nonrestrictive relative clauses. The former are simply juxtaposed to the noun they qualify while the latter are coordinated by *oo*:

Nínkú Soomáaliya ká yimí
 'The man who came from Somalia ...'
Nínkú oo Soomáaliya ká yimí
 'The man, who came from Somalia, ...'

Complement clauses are introduced by the particle *in*:

In-uu imanayo ay-aan ogahay
 Comp-3sm come-SUB FOC-1sg know-1sg
 'I know that he is coming'

Adverbial clauses expressing temporal, local, and causal circumstances are formed with relative clauses to a noun like *marka* 'time'.

Markii aan casheyayay
 Time-DET 1sg dining-PROG-PAST-1s
saaxiibkay baa soo galay
 friend-POSS1sg FOC DIR come:in
 'When I was dining my friend came in.'

Bibliography

- Abraham R C (1964). *Somali-English dictionary*. London: University of London Press.
- Bechhaus-Gerst M & Serzisko F (eds.) (1988). *Cushitic-Omoti-Papers from the International Symposium on Cushitic and Omotic Languages*. Hamburg: Buske.
- Berchem J (1991). *Referenzgrammatik des Somali*. Cologne: OMIMEE Intercultural Publishers.
- Cardona G R & Agostini F (eds.) (1981). 'Fonologia e lessico.' *Studi Somali* 1, Mimistero degli Affari esteri-dipartimento per la Cooperazione allo Sviluppo
- Comitato tecnico linguistico per L'universita' Nazionale Somalia.
- Hyman L (1981). 'Tonal accent in Somali.' *Studies in African Linguistics* 12(2), 160–202.
- Labahn T (ed.) (1983). *Proceedings of the Second International Congress of Somali Studies, vol. 1: linguistics and literature*. Hamburg: Buske Verlag.
- Lamberti M (1986). *Die Somali-Dialekte*. Hamburg: Buske Verlag (= Kuschitische Sprachstudien 5).
- Lecarme J (1991). 'Focus in Somali: syntax and interpretation; Focus en somali: syntaxe et interprétation.' *Linguistique Africaine* 7, 33–63.
- Lecarme J (2002). 'Gender "polarity": theoretical aspects of Somali nominal morphology.' In Boucher P (ed.) *Many morphologies*. Somerville: Cascadilla Press. 109–141.
- Livnat M A (1983). 'The indicator particle *baa* in Somali.' *Studies in the Linguistic Sciences* 13(1), 89–132.
- Puglielli A (ed.) (1981). 'Sintassi della lingua somala.' *Studi somali* 2, Mimistero degli Affari esteri-dipartimento per la Cooperazione allo Sviluppo Comitato tecnico linguistico per L'universita' Nazionale Somalia.
- Reinisch L (1900). *Die Somali Sprache* (3 vols). Vienna: Alfred Hölder.
- Saeed J I (1984). *The syntax of focus & topic in Somali*. Hamburg: Buske Verlag (= Kuschitische Sprachstudien 3).
- Saeed J I (1993). *Somali reference grammar*. Kensington, Maryland: Dunwoody Press. (2nd rev. edn.).
- Serzisko F (1984). *Der Ausdruck der Possessivität im Somali*. Tübingen: Gunter Narr Verlag.
- Serzisko F (1992). 'Collective and transnumeral nouns in Somali.' In Adam H M & Gesheker C L (eds.). *Proceedings of the First International Congress of Somali Studies [Held in 1980]*. Atlanta, Georgia: Scholars Press. 513–525.
- Svolaccia M, Mereu L & Puglielli A (1995). 'Aspects of discourse configurationality in Somali.' In Kiss K (ed.) *Discourse configurational languages*. New York: Oxford University Press. 65–98.
- Tosco M (2004). 'Between zero and nothing. Transitivity and noun incorporation in Somali.' *Studies in Language* 28(1), 83–104.

Songhay Languages

G J Dimmendaal, University of Cologne, Cologne, Germany

© 2006 Elsevier Ltd. All rights reserved.

The name Songai (also Songhay, Songhai, Sonrai) refers to a range of lects spoken mainly along the Niger River in Mali and Niger, as well as in Burkina Faso, and centering around major towns in the area. There are three major varieties: Western Songai (which includes Koyra Chiini, the town language of

Timbuktu, and Djenne Chiini, spoken in Djenné), Central Songai (which includes Humburi Senni, with Hombori as the major city, and Kaado), and Eastern Songai (with Koyraboro Senni as a major lect) with Gao as a major urban centre. The Gao variety has been designated as the standard for Songai in Mali. The total number of speakers in these countries is estimated to be at least 1.1 million. Zarma (Dyerma), which is spoken by some 2 million people mainly in Niger and Nigeria, and Dendi, with around 72 000 speakers mainly in Niger and Benin,

are closely related, but constitute separate languages. In addition, there are varieties in Mali and Algeria whose grammatical structure is similar to Songai, but whose lexical structure is rather deviant. Their speakers, who are culturally Tuareg, are known under a variety of names, e.g., as Tasawaq or Tadaksahak in Mali, and Korandjé in Algeria.

According to Greenberg (1963), Songai constitutes one of the six primary branches of the Nilo-Saharan phylum. Nicolai (1990) has argued that Songai is nongenetic in origin, with a Tuareg (Berber) variety playing a major lexifying role. The documentation of the Songai cluster has improved dramatically as a result of a series of monographs by Nicolai (1981) and Heath (1998, 1999a, 1999b).

The spreading of the Songai lects probably is related to the expansion of the Song(h)ai Empire from the 9th century until the late Middle Ages. Areal contact with neighboring languages belonging to different language families, such as Mande, Kwa, Gur (all Niger-Congo), and Berber (Afroasiatic) appears to have resulted in considerable typological variation within this cluster. Thus, whereas Central Songai varieties such as Humburi Senni or Kaado are tonal, western varieties such as Koyra Chiini appear to be nontonal. Also, in Western Songai varieties, SVO order appears to be common, with markers for mood, aspect, and negation occurring between the subject and the verb, and with complements other than the object following the verb. However, the object precedes the verb in Central and Eastern Songai varieties, which also use a transitive marker before the object noun phrase. All Songai lects

appear to use postpositions. Nominal modifiers tend to follow the head noun, but possessors precede the latter. The use of a one-term deictic marker appears to be a more common areal phenomenon, also attested in neighboring Mande languages. Affixational morphology is somewhat restricted (derivational morphology in the verb, for example, appears to involve mainly causative and centripetal marking), but cliticization of morphemes is highly common in Songai, frequently resulting in a mismatch between phonological and grammatical words. Logophoric marking, as a reference tracking mechanism and evidential hedging strategy, is also used across sentence boundaries; compare Heath (1999a: 322–328).

Bibliography

- Greenberg J H (1963). 'The languages of Africa.' *International Journal of American Linguistics* 29(1,2).
- Heath J (1998). *Wortkunst und Dokumentartexte in afrikanischen Sprachen 6: Texts in Koroboro Senni: Songhay of Gao, Mali*. Cologne: Rüdiger Köppe.
- Heath J (1999a). *Mouton Grammar Library 19: A grammar of Koyra Chiini: the Songay of Timbuktu*. Berlin/Philadelphia: Mouton de Gruyter.
- Heath J (1999b). *Westafrikanische Studien 19: A grammar of Koyraboro (Koroboro) Senni: the Songhay of Gao, Mali*. Cologne: Rüdiger Köppe.
- Nicolai R (1981). *Les dialectes du songhay*. Paris: SELAF.
- Nicolai R (1990). *Parentés linguistiques (à propos du songhay)*. Paris: Centre National de la Recherche Scientifique.

Sorbian

G H Toops, Wichita State University, Wichita, KS, USA

© 2006 Elsevier Ltd. All rights reserved.

Sorbian is one of three branches of the West Slavic languages comprising Lower Sorbian and Upper Sorbian (the other two branches being the Czech-Slovak and the Lechitic). 'Sorbian' thus serves as a convenient cover term for one or both of the Sorbian literary languages and their respective dialects. The Lower Sorbian (hereafter, LSo) and the Upper Sorbian (USo) literary languages constitute supradialectal norms generally used in writing and public or mass communication (print media, radio, and television); in informal settings Sorbs (both Lower and Upper) tend to speak the dialect characteristic of their native

village, with occasional admixtures of literary elements and German vocabulary. Sorbian as defined here is spoken today as a native language entirely within the borders of the Federal Republic of Germany (from 1949 to 1990, within the borders of the former German Democratic Republic); more precisely, it is spoken completely within the eastern German region of Lusatia (LSo *Łužyca*, USo *Łužica*, German *die Lausitz*), situated partly in the German state of Saxony (*Freistaat Sachsen*) and partly in the state of Brandenburg (see **Figure 1**). The Brandenburg portion includes what is traditionally known as *Lower Lusatia* (German *Niederlausitz*), while the Saxon portion includes most of *Upper Lusatia* (German *Oberlausitz*). These geographic designations are sometimes applied to the languages spoken there, whence

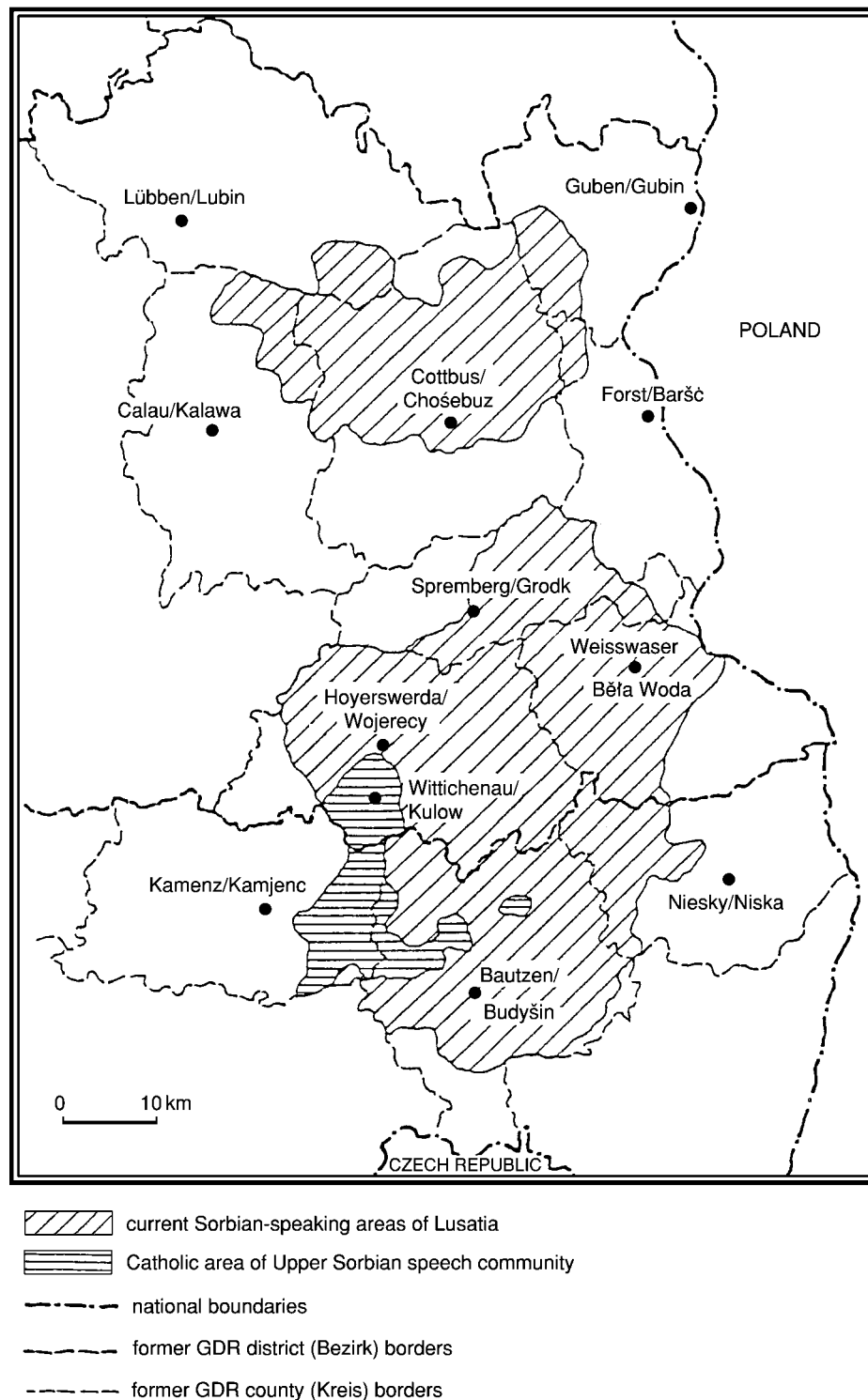


Figure 1 Sorbian speech communities. Schiller K J and Thiemann M (1979), *Stawizny Serbow* (vol. 4), Bautzen: Domowina, with permission.

the terms 'Lower Lusatian' and 'Upper Lusatian' in lieu of 'Lower Sorbian' and 'Upper Sorbian,' respectively. The Sorbian-language area, like the Sorbian speech community itself, has shrunk considerably in the past

100–150 years, so that it now extends at most only about 90 kilometers north-south and some 55 kilometers from west to east inside Lusatia proper. The Lower Sorbs call themselves *Serby* in their own

language but reveal a preference for the appellation *Wenden* 'Wends' in German; the Upper Sorbs call themselves *Serbja* (*Sorben* [or, more specifically, *Obersorben*] in German). Descendants of Sorbs (now all English-speaking) who settled in the American state of Texas (some 60 kilometers east of the state's capital city, Austin) in 1854 describe their heritage as 'Wendish.' As far as anyone has determined, all Lusatian Sorbs – to the extent that they still speak some form of Sorbian – are bilingual in Sorbian and German and have been so since the early decades of the 20th century. Partly as a result of this universal bilingualism, ethnic Sorbs have tended increasingly to become unilingual German-speakers.

Estimates of the current number of native Sorbian speakers vary. An ethnosociological poll conducted by the Institute of Sorbian Ethnography (*Institut za serbski ludospyt*) in 1987 suggested 67 000 as the maximum number of Sorbian speakers in both Lower and Upper Lusatia (Faska, 1998: 20). A more recent compendium of USo grammar puts the number of USo speakers at no more than 53 600 (Schaarschmidt, 2002). Assuming that such estimates are accurate, one is led to surmise a maximum of 13 400 speakers of LSo (67 000 minus 53 600), a figure that is not substantially at odds with the estimate of 16 000 LSo speakers cited by Šatava on the basis of data also collected in 1987 (Šatava, 1994: 198). The overwhelming majority of today's native LSo speakers are more than 60 years old; consequently, the LSo dialects are expected to be extinct within the next 15–25 years (Jodlbauer *et al.*, 2001: 204). Schaarschmidt reckons that USo should be extinct by the year 2070; however, given current efforts at language maintenance, the USo literary language and at least some of its dialects (notably the so-called 'Kamjenc' [German: Kamenitz] dialect of the approximately 15 000 Catholic Sorbs northwest of Budyšin [Bautzen]) stand a good chance of surviving well beyond the year 2100 (Schaarschmidt, 2002: 5–6).

The center of LSo literary and cultural activity (including radio and television broadcasting) is the city of Chošebuz (Cottbus); the center of USo literary and cultural activity is the city of Budyšin (Bautzen). This reflects the emergence of these two cities as 'dialect centers' in the 17th–18th centuries, owing to the fact that: (a) the majority of those educated Sorbs who translated (mostly religious) texts from German into Sorbian hailed from, or from the vicinity of, these cities; and (b) these cities already occupied major economic and political positions in Lusatia at that time. The first cohesive text written in Sorbian that we know of is the so-called 'Budyšin Oath' (*Budyska přisaha*) or 'Wendish Citizen's Oath' (*Bürgerleid Wendisch*) dating from the year 1532. The early production of longer Sorbian texts is connected with the

spread of the Protestant Reformation in Lusatia and the ensuing Thirty Years' War (1618–48). The first Sorbian religious text, as far as we know, is an eastern LSo translation of the New Testament of Martin Luther's Bible, which was written by hand in 1548. The first Sorbian printed book is a LSo collection of church hymns and a small Lutheran catechism published by the preacher Albin Moller in 1574. The first printed book in USo is a translation of the Lutheran catechism published by the preacher Wenzeslaus Warichius in 1587. Such 16th-century texts already exhibited varying degrees of German lexical and grammatical influences (e.g., Sorbian use of the demonstrative pronoun/adjective as a reflection of the German definite article).

Because Protestantism did not completely take hold among the Upper Sorbs, Budyšin emerged as a dialect center only for the USo Protestants; among the USo Catholics, the dialect spoken in and around Kulow (Wittichenau) emerged in the 17th century as the basis for a Catholic variant of a nascent USo literary language. The dialectal basis for this variant eventually broadened in the direction of a West Sorbian Catholic dialect situated in the vicinity of Chrósćicy (Crostwitz). Thus, by the 18th century, two literary languages – one of them with two variants – existed among the Sorbs: LSo, Protestant USo, and Catholic USo.

Like German books printed at the time, the earliest Sorbian publications were printed in German black-letter, or Fraktur. Sorbian spelling was based on sound correspondences with German graphemes or phonetic approximations of them. Thus, the German trigraph *fch* might correspond to the graphemes *ř*, *š*, or even *ž* (representing palatal continuants) in today's USo orthography. The palatal affricates, in contrast, were graphically influenced to some extent by Polish – *cz* was used where today one finds USo *č* or *ć*. For example, in Warichius's catechism of 1597, we find *The džesacz kśfni Bohšche* (in contemporary USo orthography: *Te džesać křŋni Bože*) 'God's ten commandments' (cf. Schuster-Šewc, 1967: 52). Around the middle of the 19th century, efforts were made to reconcile the Catholic and the Protestant variants of literary USo by standardizing their orthographies. The resulting orthography, set forth in the 1848 publication of *Hornjolužiski serbski prawopis z krótkim rěčničnym přehladom* ('Upper Lusatian Sorbian orthography with a brief grammatical overview') by Christian Traugott Pfuhl (Křesćan Bohuwěr Pful), incorporated Czech and Polish conventions – use of diacritics like the Czech háček (*č*, *ě*, *ř*, *š*, *ž*) and the Polish acute accent (*ć*, *dź*, *ń*, *ó*) as well as the Polish velar *l* (*ł*) – and was therefore labeled 'analogical.' The new orthography, however, was

also etymologically based, introducing graphemes – particularly in word-initial consonant clusters – that had no phonetic value: *łowa* → *hłowa* ‘head,’ *cyć* → *chcyć* ‘to want,’ *dže* → *hdže* ‘where,’ *zac* → *wzac* ‘to take.’ On the one hand, the etymologically based orthography has since led to a number of artificial spelling pronunciations; on the other hand, it has made written USo more readily interpretable for those familiar with other Slavic languages. Moreover, it disambiguates a large number of homophones – e.g., *wóz* ‘wagon, car,’ *łós* ‘elk,’ *hlós* ‘voice,’ and *włós* ‘hair’ – all of which are pronounced [uú’s]. LSo remains less reflective of etymology, cf. LSo *cu* ‘I want,’ *žo* ‘there,’ *ned* ‘right away,’ and *cora* ‘yesterday’ vs. USo *chcu*, *hdže*, *hnydom*, and *wčera*, respectively. An USo spelling reform was introduced again in 1948. Several USo orthographic conventions have been adopted for LSo (*inter alia*, representing palatalization by means of the letter *j* rather than an acute accent—*młod* → *mjod* ‘honey,’ *ńasć* → *njasć* ‘to carry’; cf. Schuster-Šewc, 1996: 253 and 260). LSo influence on USo spelling can be seen in the substitution of word-initial *ch* for earlier *kh* (USo *khodźić* → *chodźić* ‘to go’; LSo *chójźis* ‘idem’).

Today’s USo alphabet consists of the following graphemes: *a, b, c, č, d, dž, e, ě, f, g, h, ch, i, j, k, l, ḷ, m, n, ŋ, o, ó, p, r, ř, s, š, t, ć, u, w, y, z, ž*. The letters *v* and *x* occur in foreign names. The letters *č* and *c’* represent the same phoneme /tʃ/ while reflecting different etymologies; the same holds true for the letters *l* and *w* (phonetically [u]). The letter *ř* occurs only after *k, p*, and *t* and is pronounced like *š* except where *tř* constitutes a digraph representing the phoneme /c’/ (e.g., *tři* /c’i/ ‘three’). Today’s LSo alphabet includes: *a, b, c, č, ć, d, e, ě, f, g, h, ch, i, j, k, l, ḷ, m, n, ŋ, o, p, r, ř, s, š, ś, t, u, w, y, z, ž, ź*. Unlike USo, LSo alphabetizes *ch* with *c*, rather than after *h*. In 1995, the LSo Language Commission relegated *ó* (phonetically [e] or [y] after labials and velars) to language-teaching materials, replacing it with *o* (Starosta, 1999: 19).

Both LSo and USo exhibit grammatical features that set them apart from other contemporary Slavic languages. The LSo verb paradigm still includes the supine found in early Slavic. USo retains a rich system of tenses – present, preterite (also called ‘aorist’ if the verb is aspectually perfective, ‘imperfect’ if it is imperfective), future, perfect, and pluperfect; in addition, the literary language and a number of USo dialects retain the iterative preterite tense (formally identical to the conditional mood). The LSo dialects exhibit only one past tense, formed with the auxiliary *bys* ‘to be’ and the *l*-participle; literary LSo, in contrast, exhibits all the tenses of USo, artificially

(re)created and phonologically adapted. The Sorbian grammatical category of number is expressed as singular, dual, and plural, which is marked both on the noun or pronoun and on the verb. The dual is gradually giving way to the plural in the dialects, often surviving only exceptionally after the quantifier ‘two’ (USo *dwaj, dvě*) with plural agreement in the verb (*Dvě knize su na blidže leželi* in lieu of literary USo *Dvě knize stej na blidže ležaloj* ‘Two books lay on the table’). Nouns (substantive and adjective) and pronouns exhibit six cases – nominative, genitive, dative, accusative, instrumental, and locative. A vocative form exists in USo, but only for masculine nouns (*Šćěpan* → *Šćěpano! Šćěpanje!*) and one feminine noun – *mać* (*maći!*) ‘mother.’

Sorbian word order is basically SOV (Subject-Object-Verb); however, compound verb tenses and the clitic status of the auxiliary verbs usually produce a ‘bracket construction’ like that found in German (so-called *Rahmenkonstruktion*). Sorbian subordinate clauses, in contrast, generally do not imitate the German clause-final placement of finite verbs.

The LSo and the USo dialects are connected by a zone of ‘transitional’ dialects. Here, LSo lexical and morphophonological features increase and USo ones decrease as one moves from south to north, while USo features increase and LSo ones decrease as one moves from north to south. Although this might suggest a single dialectal continuum (hence, a single Sorbian language), in fact, the degree of mutual intelligibility between LSo and USo proper is perceptibly less than that which exists today between Czech and Slovak.

Bibliography

- Faska H (ed.) (1998). *Serbščina*. Opole: Uniwersytet Opolski. (Najnowsze dzieje języków słowiańskich.)
- Jodlbauer R, Spieß G & Steenwijk H (2001). *Die aktuelle Situation der niedersorbischen Sprache. Ergebnisse einer soziolinguistischen Untersuchung der Jahre 1993–1995*. Bautzen: Domowina-Verlag.
- Šatava L (1994). *Národní menšiny v Evropě. Encyklopedická příručka*. Praha: Ivo Železný.
- Schaarschmidt G (2002). *Upper Sorbian*. Munich: LINCOM Europa.
- Schuster-Šewc H (1967). *Sorbische Sprachdenkmäler: 16–18 Jahrhundert*. Bautzen: Domowina-Verlag. (Spisy Instituta za serbski ludospyt, 31.)
- Schuster-Šewc H (1996). *Grammar of the Upper Sorbian language: Phonology and morphology*. Toops G H (trans.). Munich and Newcastle: LINCOM Europa.
- Starosta M (1999). *Dolnosorbsko-nimski słownik/Niedersorbisch-deutsches Wörterbuch*. Budyšin: Domowina-Verlag.

South Asia as a Linguistic Area

K Ebert, University of Zurich, Zurich, Switzerland

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The roughly 450 languages of South Asia belong to four different language families: Indo-European, Dravidian, Austroasiatic, and Sino-Tibetan. There are three small isolates, Burushaski, Nahali, and Kusunda (probably extinct). Speakers of Indo-Aryan (IA) languages constitute 78% of the inhabitants of South Asia, followed by Dravidian (DRAV) speakers with 20%; speakers of Austroasiatic, i.e., Khasi and Munda (MU), and Tibeto-Burman (TB) languages together do not constitute more than 2%. The number of speakers is reflected in the space devoted to languages in the sprachbund literature. Emeneau, the first authority on the subject, wrote about 'India as a linguistic area' (1956). Although Nepal and Bhutan also belong to South Asia, languages of the Himalayas are seldom included in the sprachbund literature, nor are languages of Nagaland or Meghalaya.

I shall first look at the features most often mentioned as characterizing the area. After a brief summary of the history of the field, I shall discuss the possibility of interpreting the linguistic data as evidence for earlier settlement patterns and migrations. The last section stresses the need for more detailed investigation of subareas.

Areal Features

The following features are mentioned in most of the literature on the South Asian sprachbund:

- retroflex consonants
- OV word order
- converbs ('conjunctive participles')
- compound verbs
- the quotative
- morphological causatives
- dative subjects.

Retroflex Consonants

All the major languages of South Asia have a phonemic opposition between retroflex and dental consonants (Ramanujan and Masica, 1969). Many IA and most DRAV languages also have retroflex *ɽ*, *ɳ*, and *ʅ*. There are no retroflexes in Assamese and in South MU nor in Indo-European. Kuiper's (1967) demonstration of the increasing frequency of retroflex

consonants in the successive books of the *Rigveda* is generally accepted as proof of DRAV substratum influence.

OV Word Order

The second feature, OV word order, holds for languages from Pakistan to Assam and from Nepal to Sri Lanka. The only two exceptions are Kashmiri (IA) and Khasi (AA), which are both VO. However, OV word order also characterizes the languages of Central and Northeast Asia, including Korean and Japanese.

Converbs

This feature has been extensively discussed in the sprachbund literature. But although most languages of the subcontinent have at least one converb (conjunctive participle, gerund), there are considerable differences in detail. The most general or sequential converb can fulfill a number of functions, depending on the existence of more specific converbs. Hindi has only one form (*-kar/-ke*), which has a broad range of applications. The following examples illustrate sequential, modifying, and causal interpretations of the Hindi and Tamil general converbs.

(1) HINDI (IA)

- a. us-ne nahaa-**kar** khaanaa khaa-yaa
he-ERG bathe-CONV meal eat-PFV:3SG:MASC
'Having bathed he ate his meal'
- b. vah dhaur-**kar** aa-yaa
he run-CONV come-PFV:3SG:MASC
'He came running'
- c. vah raat din kaam kar-**ke** biimar
he night day work do-CONV ill
par ga-yaa
fall GO-PFV:3SG:MASC
'He fell ill because he worked day and night'

(2) TAMIL (DRAV)

- a. avan inkee va-**ntu** enn-ai.k kuuppitt-aan
he here come-CONV I-ACC call:PT-3SG:MASC
'He came here and called me'
- b. avan oot-i va-nt-aan
he run-CONV come-PT-3SG:MASC
'He came running'
- c. mazai peytu kulam nirai-yum
rain fall:CONV pond fill-FUT:3SG:NEUT
'It rained and (therefore) the pond will fill'

Most languages have a simultaneous or modifying converb different from the conjunctive participle, which is often reduplicated. See the following examples with Hindi (1b), and Tamil (2b).

- (3) ORIYA (IA) baṭore ja-u ja-u
 'walking along the road'
MALTO (DR) eek-no eek-no paawno
 'walking along the road'
ATHPARE (TB) huk-sa huk-sa abe
 'he came down barking'

(The sequential converbs are Nepali *-ii*, *-era*, Oriya *-i*, *-iki*, Malto *-ka/* + person markers, Athpare *-ung* after finite markers.) The eastern IA languages (Assamese, Bengali, Oriya) moreover have a conditional converb, as do most Dravidian languages.

- (4) ASSAMESE (IA) moi aahi-le 'if I come'
ORIYA (IA) se asi-le 'if he comes'
TAMIL (DR) mažai peyt-aal 'if it rains'
KANNADA (DR) avaru oodid-are 'if he studies'

Santali has nonfinite forms in *-kate* and *-te* (5a); the quasi-converbs carry finite tense-aspect and person markers, though they lack the finite marker *-a* (5b).

- (5) SANTALI (MU)
a. calak'-calak'-te mir'-ṭaṇ
go:MID-REDPL-CONV.SIM one-CLASS
toyo-ko jæl-tiok'-ked-e-a
jackal-S:3PL see-reach-PT-O:3SG-FIN
'While they were walking, they got sight of a jackal'
b. kəṭeč'-ked-e-khon ḍaṅgra-də-e
castrate-PT-O:3SG-ABL bullock-TOP-S:3SG
luṭkum-en-a
fat-PT:MID-FIN
'Since they castrated it, the bullock became fat'

The South Asian picture is far from uniform, and the number of converbs can vary from one to ten or more (e.g., Hayu), although three or four is the norm. Converbs are even more characteristic of Central and Northeast Asian languages, where converbal forms mark all types of adverbial subordination. Moreover, converbs seem to go together with OV word order (Ethiopian Semitic and Quechua). So do adnominal 'relative participles,' which are sometimes mentioned together with converbs as an areal trait of South Asia.

Compound Verbs

South Asian languages form compound verbs consisting of the general/sequential converb form of the main verb (V1) followed by a finite form of a second verb (V2). The second verb, which also occurs as a full verb, is semantically bleached, but not fully grammaticized. The inventory of second verbs listed in grammars differs somewhat from language to language, and – as the list is not closed – also from author to author. The most frequent V2s include:

1. directionals with the full verb meanings 'go', 'come';
2. disposals that express that something is done away with, such as 'throw', 'send', 'put aside';
3. verbs that express the suddenness or unexpectedness of an event, such as 'fall', 'rise';
4. 'give' and 'take' have auto- and other-benefactive meanings as V2.

- (6) MARATHI (IA)
sagle kaagad mi phaar-un ṭaak-le
all paper I tear-CONV V2:THROW-PT
'I tore up all the papers'

BENGALI (IA)
se mar-e gela
he die-CONV V2:GO:PT
'He died'

KANNADA (DR)
avaru sattū-hoo-daru
s/he die:CONV-V2:GO-PT:3PL(HON)
'He died'

The terminology used for this construction is rather inconsistent. Apart from a plethora of names for the V2 (explicator, vector, aspectivizer, light verb), some authors use the term *serial verb* instead of compound verb. Sometimes constructions with phasal verbs (which are not semantically bleached and often combine with the infinitive) and grammaticized forms are included.

As with converbs, a closer look at compound verbs yields a rather diverse picture. First, the shape often differs from the canonical 'V1-CONV + V2 finite' pattern. Some languages have developed new converb suffixes or an optional long form. The new or long forms are used in clause combining, but not in compound verbs. Thus Hindi and Panjabi never have the converbal suffixes *-kar* or *-ke* in compound verbs, but use an old form now reduced to the bare stem of V1 (see *par ga-yaa* in (1c)); Oriya never has *-iki* or *-ikori*; Kōḍava never has the converb in *-iti*, but the old converb, which is identical with the past stem (see the Tamil and Kannada examples above).

- (7) ORIYA (IA)
pila-mane pōrh-iki ghōr-u
child-PL study-CONV house-ABL
bahar-i-gol-e
go.out-CONV-V2:GO:PT-3:PL
'After studying the children went out of the house'
KODAVA (DR)
ava seebi tind-iti
she apple eat:PT-CONV
catti-pooc-i
die:PT:CONV-V2:GO:PT-3
'She ate the apple and died'

Santali root compounds look much like the Hindi and Panjabi forms. Languages that make little use of converbs often have finite markers on both verbs, such as Kurukh and some MU and TB languages.

- (8) SANTALI (MU) *təl-uric'-ked-e-a-e*
 tie-V2:FIRM-
 PT-O:3SG-FIN-S:3SG
 'he tied him up firmly'
 KURUKH (DR) *iirk-an-cicck-an*
 see:PT-1SG-V2:GIVE:PT-1SG
 'I looked after it'
 PARENGI (MU) *silay-ing-ta'y-ing*
 sew-1SG-V2:GIVE-1SG
 'sew for me!'
 CAMLING (TB) *c-ung-pak-ung-a*
 eat-1SG-V2:PUT-1SG-PT
 'I ate it up'

Second, the inventory and semantics of second verbs sometimes differ in substantial ways from more canonical patterns, especially in MU and N-DRAV. The unusual V2s include Santali *jaora* 'gather', *uric'* 'make firm', *eset'* 'close', *nyam* 'find', *anga* 'dawn', and Kurukh *xacc-* 'break', *biɁ-* 'cook'. Tamil also has unusual second verbs, and even the common ones show irregular semantics, often conveying purely emotional meanings.

Compound verbs of the form "V1-CONV + V2 finite" are typical of Altaic (with V2 called postverb, descriptive verb, or auxiliary). Turkic and Mongolic languages share the second verbs mentioned under (a) to (d) above. But unlike South Asian languages they make regular use of postural verbs as atelicizers or durative markers. Otherwise the use of second verbs in languages in the northern part of India appears to be more similar to Turkic-Mongolian than to South Dravidian, though further detailed studies are needed.

Quotatives

Quotatives of the form 'say'-CONV, such as Bengali (IA) *bole*, Nepali (IA) *bhanera*, Telugu (DR) *ani*, Santali (MU) *mente*, and Sora (MU) *gamle* correspond to Uzbek (TURKIC) *deb*, Mongolian *geɣ* (and to Ethiopic Inor *barə* Quechua *nishpa*). There is nothing remarkable about the development into complementizers, which are then used together with verbs characterizing speech or mental acts, or with onomatopoeic words. The development from 'say' to complementizer is also attested in African languages and Creoles. It is an ongoing process that can be observed in South Asia, and not all languages are at the same stage. The last stage, comparatives marked by the quotative, is attested only in some languages of Nepal (e.g., Newar, Nepali) and South Dravidian. Further detailed

studies are necessary to describe the use of quotatives in the single languages and to distinguish borrowings from internal developments.

Morphological Causatives

All languages of South Asia show morphological causatives, and most have secondary or indirect forms.

- (9) HINDI (IA) *siikh- / sikhaa- / sikhwa-*
 'learn / teach / have taught'
 KASHMIRI (IA) *con- / caavun- / caavinaavun-*
 'drink / give to drink / let give to drink'
 MALAYAL (DR) *oti- / otikk- / otippik-*
 'break (intr/tr) / let break'

The patterns differ toward the east, where prefixes prevail. AA languages thus conform to their relatives in Southeast Asia: Khasi *tip / pn-tip* 'know / inform', *iap / pn-iap* 'die / kill', and Sora *jum / a-jumjum* 'eat / feed'. TB languages of the east also have prefixes, e.g., Mao Naga *apo / so-pho* 'break (intr/tr)'; Mikir *thi* 'die', *pe-thi* 'kill', *pa-pe-thi* 'let kill'. In some languages, double causatives can receive a simple causative interpretation; Kharia (MU) *doko* 'sit down', *ob-do-b-ko-yo?* (CAUS-sit-CAUS-sit-PTII) 'he made him make her sit' or 'he made him sit' (Zide and Anderson, 2001: 521, 523).

Morphological causatives, including double causatives, are also found in the languages of Central and Northeast Asia.

Dative Subjects

This construction is familiar from some European languages such as Latin and German. An experiencer is coded by the dative (in Eastern IA languages by the genitive), and the dative constituent behaves like a subject in some respects (Verma and Mohanan, 1990).

- (10) HINDI (IA) *bacce-ko thaṇḍ lag rahii hai.*
 child-DAT cold feel PROG:FEM:SG is
 'The child is/feels cold'
 TAMIL (DR) *ena-kku kuḷiraa iru-kk-utu.*
 he-DAT cold be-PRES-3SG:NEUT
 'He is/feels cold'

The construction is less prominent in MU and TB; some languages even lack a dative or an oblique case marker. Nevertheless, it counts as a strong areal feature; as in contrast to the constructions mentioned above, it does not exist in Altaic languages. As it is not typical of Sanskrit, DRAV is usually considered a likely source.

Other Features

Several other features were claimed to be characteristic of South Asia, but most were dropped later, e.g., aspirated consonants, nasalized vowels, classifiers,

ergative constructions, no prefixes, no verb for 'have'. The proposed features were evaluated by Masica (1976: 187–190); most of them turned out to be irrelevant or of little relevance for the South Asian linguistic area. Classifiers and echo formations were suggested in Emeneau's pioneering article 'India as a linguistic area' (1956). Emeneau erroneously considered classifiers, which occur especially in eastern and central India, as borrowings from IA into the individual MU and DRAV languages. But classifiers clearly have swept over from Southeast Asia. Emeneau's conclusion probably has to be ascribed to the fact that the classifiers of MU and N-DRAV languages often have an IA shape; i.e., the morphemes are borrowed, but the function is not.

Most South Asian languages have a construction in which a word is reduplicated, replacing the first consonant (sometimes also the following vowel), so that the second word constitutes an echo of the first one. The echo expresses 'and such things', e.g., Tamil *kudirai-gidirai* 'horses and such things'. The consonants used in echo formation are distributed in areal patterns (see map in Trivedi, 1990: 80–81). Dravidian languages of the south show exclusively *k*, *g*; Orissa and Bengal have *p*, *ph* or *t*. The patterns are independent of genetic affiliation; compare Oriya (IA) *iskul-phiskul* 'school and such', *cini-phini* 'sugar and such' with Ho (MU), *ɔpis-pɔpis* 'office and such'; Bengali (IA) *ghusur-tusur* 'pigs and such' with Nocte (TB) *san-tan* 'sun and such', and Santali (MU) *bəkɔp-təkɔp* 'brothers and such'. Emeneau considered this feature to be borrowed from DRAV, as he took it to be otherwise unknown in Indo-European. (It is rare in IE, though not in Altaic; Turkish *kitap kitap* 'books and such things', Uzbek *nān pān* 'bread and other baked goods').

A more promising candidate seemed to be the Sanskrit particle *api*, corresponding to Tamil *-um*, which has five functions reconstructable for Proto-Dravidian: 1. additive focus, 'also', 2. 'and', 3. 'even', 4. totality with numerals ('every'), 5. together with question words it yields indefinite pronouns. Emeneau found the five functions in all subgroups of DRAV, but only in a few modern IA languages (e.g., Marathi, Oriya). As not all functions exist in early Vedic, he concluded that the functions of Sanskrit *api* developed by analogy with the DRAV model (Emeneau, 1980: 218). No parallels in MU or TB are mentioned. Nevertheless, it is one of the few features that Masica considers to be area-defining, though the criteria are not further discussed and remain unclear. The bundle of functions 1.5. is far from rare for a particle meaning 'also, even', and parallels could be cited from Altaic and many other languages.

History of the Field

South Asia counts as a classic example of a linguistic area. As early as the 19th century, Indologists noticed some common traits between IA and DRAV. The discussion centered around the question of whether IA could have adopted certain features from DRAV, e.g., retroflex consonants or converbs. Sanskritists have at all times tried to minimize such possible influence and proposed internal developments (Hock, 2001). Emeneau, doing intensive research on DRAV, was the first to put the comparison on a more solid base. In his early article 'India as a linguistic area,' he came to the conclusion that "the languages of the two families, Indo-Aryan and Dravidian, seem in many respects more akin to one another than Indo-Aryan does to other Indo-European languages" (1980: 119–120; 1956). His definition of the term *linguistic area* as "an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families" is still useful, in contrast to his later proposals. In the introductory article to the reprint volume (Emeneau, 1980), he critically reviews his own earlier work and adds the conditions: For a feature to be area-defining it has to be "pan-Indic and not extra-Indic" (1980: 2). Only if several such features are found and the area is delimited by a bundle of near isoglosses can the linguistic area, in his view, be considered established. A second step has to show the origin of the areal features and their distribution. If a feature can be reconstructed for language X, it must have been borrowed into language Y.

The newly formulated conditions turn out to be problematic. If a South Asian areal trait must not be extra-Indic, most of the proposed features have to be abandoned. But this criterion does not seem to have been taken too seriously anyway; few researchers have bothered to look outside the borders of South Asia. This is true even of Emeneau himself, as evidenced by his proposals for classifiers and echo words as areal traits. And the criterion of pan-Indianness was not tested. Areal relevance was mostly claimed on the basis of a few examples from IA and DRAV, sometimes adding one or two examples from a MU language. Isogloss bundles were not shown, as Emeneau remarks: "Unfortunately I know of no demonstration of such a bundling of isoglosses" (1980: 128).

In an important article, Kuiper (1967) examined Vedic and Sanskrit texts that could shed light on the origin of the South Asian linguistic area. He investigated the appearance of retroflex consonants, converbs ('gerunds'), and the development of

the quotative *iti* in IA. Regarding retroflexes, he comes to the conclusion that prehistoric bilingual speakers of IA – presumably native speakers of a Dravidian language – reinterpreted IA allophones in terms of their native system, thus establishing a novel phonemic distinction in IA. Kuiper further traces the gradual increase of converbs in the successive *Rigveda* texts and ascribes the development again to bilinguals, who would have used converbal constructions first in colloquial speech, from where it crept into more formal registers. The Sanskrit particle *iti* ‘thus’ originally introduced quotations and is attested in initial position in Vedic texts. Gradually it became post-quotative by analogy with the Dravidian ‘say’-CONV, and in later Sanskrit this was the standard.

A further influential article was Southworth’s contribution to the volume edited by himself and Apte (1974). Southworth found that the frequencies of retroflex consonants in modern IA languages decreases from west to east. Western IA languages such as Marathi, Gujarati, and Panjabi show a ratio of 3:1 for dentals and retroflexes, which corresponds to the ratio in DRAV, but Bengali has 12:1 (1974: 212). Southworth interprets this, together with gender marking, as evidence for a DRAV substratum in the west; the lower number of retroflexes and the presence of classifiers is interpreted as a reflex of a TB substratum in the Ganges delta and the east. The assumption that the distribution of features today mirrors the situation that obtained 2000 years ago is of course problematic.

The first comprehensive and systematic study on some features in South Asia and beyond, Masica’s *Defining a linguistic area: South Asia* (1976), has become a standard text. Whereas earlier publications on the sprachbund were confined to demonstrating shared features of South Asian languages (if not just IA and DRAV), Masica’s concern was to find out to what extent these features are purely South Asian. The results are rather devastating for the sprachbund hypothesis if based on the conditions formulated by Emeneau: of the five traits investigated, only one, namely dative subjects, turned out to be specific for South Asia. The other four – morphological causatives, OV word order, converbs, and compound verbs – are equally characteristic of most languages of Central and Northeast Asia, as already mentioned. Researchers have since also spoken of an Indo-Turanian area.

The idea of a South Asian sprachbund is thereby not invalidated. The clustering of the two essential features, dative subject and retroflexes, together with shared idioms and semantics (Emeneau, 1980: 236, 250; Masica, 2001: 258) and the OV characteristics,

some of which demonstrably spread from South Asian centers, together add up to the often perceived ‘Indianness’ of South Asian languages.

Historical Evidence

One of the aims of identifying linguistic areas is to find evidence for earlier settlements and migrations. Documentation of South Asian languages reaches back to the second millennium B.C. for Sanskrit and back two millennia for Tamil, but little is known about earlier stages of MU and TB. And in spite of the early attestations of IA and DRAV, the substratal influence of the latter on the former remains controversial. According to Hock (2001), IA and DRAV were typologically more similar at the time of the earliest contacts than is commonly assumed. Similarities such as the combination of finite marked forms, as still preferred in North and Central Dravidian, have been demonstrated by Steever (1993) for older stages of South Dravidian. Many of the South Asian traits could be retentions, being strengthened of course by area-specific preferences. Hock criticizes the still prevailing practice of drawing conclusions from comparisons of Sanskrit with (mainly) modern South Dravidian, ignoring older traits of Dravidian and setting aside 2000 years of history.

Even the DRAV origin of retroflexes, a seemingly solid cornerstone of the substratum hypothesis, has been questioned. This also casts doubt on the general assumption that Dravidians were the inhabitants of the Indus Valley at the time the Indo-Aryan infiltration into South Asia started. Witzel (1999) posits a ‘Para-Munda’ substratum of the oldest IA documents, i.e., the Vedic texts, which originate from the valley.

More recent influences can be traced by means of quantitative areal investigations. Hook (1987), trying to get “at the grain of history,” reports an interesting finding from a questionnaire investigation. In the returns from south of Goa, all subordinate clauses were preposed; the percentage gradually decreased toward the northwest, i.e., with the distance from the Dravidian model. West of the Indus, all clauses were postposed. This pattern is independent of the fact that OV word order typologically often goes together with preposed clauses and demonstrates the fading out of a typical feature toward the edge.

A quantitative analysis of this type can sometimes show the areal spread of a trait, though not its origin. The inference from present-day frequencies to situations 2000–3000 years back (Southworth, 1974) is hardly reliable. Hook’s endeavors to trace the possible origin of compound verbs on the basis

of frequency counts lead to no satisfying results. Today compound verbs are most abundant in the Ganges plains (Hindi 15–20% of total verbs, modern Bangla 10–13%, modern Marwari 13–18%). But comparison with the situation in 16th century Bangla (2%) and Marwari (1.5%) shows this to be a recent phenomenon (Hook, 2001: 109). If borrowed at all, the further development of the construction must be ascribed to internal forces. Hook's answer to the question of what we can conclude from the present-day distribution of the compound verb in the languages of South Asia is: "Not much." (2001: 110). Various processes of borrowing, internal developments, and even of loss are possible scenarios.

Much of the linguistic history of South Asia remains in the dark. Shifts in language have not been uncommon, and in some cases have occurred more than once, e.g., from MU to DRAV to IA. Widespread bilingualism and multilingualism have repeatedly lead to pidginization and the development of lingua francas, such as Calcutta-Hindustani or Hindi-Urdu in Bombay. Nagamese, based on IA Assamese, has become the mother tongue of the Tibeto-Burman Kachari. It is at present the only common medium of speakers of the 30 or so Tibeto-Burman Naga languages. Sadari, a Hindi based pidgin, has become the language of identification for groups who have abandoned their former DRAV or MU tongues. Some of the modern literary languages may have started out as pidgins too, as Southworth (1971) suggests for Marathi.

Historical investigations are restricted to what has been accepted into the written language. But written texts are far from an ideal basis for areal investigations, and written norms are especially conservative in South Asia. At all times, the spoken language has differed a great deal from the written one, and borrowings must have been much more abundant than texts reveal. Present investigations into languages used for everyday communication show remarkable structural convergences in some areas. The most well-known case is Kupwar, as described by Gumperz and Wilson (1971). The contact situation between Kannada, Marathi, and Urdu at the border between Maharashtra and Karnataka has lead to one single grammatical system with language-specific lexemes. This study also reveals another characteristic of linguistic areas: structural traits, especially in syntax, are largely unconscious and easily borrowed, whereas there may be considerable resistance toward lexical borrowing, especially if the language functions as a means of identification.

Subareas

Setting up isoglosses simply on the basis of the occurrence of a feature, as in Masica (1976), is apparently not an adequate technique for showing the intricate patterns of a linguistic area. In Masica's map of the dative subject, for example, the isogloss for this construction includes the northeastern corner of India. The map does not show that this feature is mainly restricted to IA and that numerous small languages of the area have not adopted it. As Masica himself notes (1976: 172), the isogloss maps do not show the gradual fading out of features. They also do not indicate different codings. If causatives are marked by suffixes in the west, but by prefixes in the east of the subcontinent, this is highly relevant for areal studies.

One of Masica's four Indo-Altaic isoglosses deviates from the others: The line for secondary causatives runs approximately along the 84th meridian, i.e., cutting off Bihar, parts of Orissa, and the northeastern provinces from the rest of South Asia. This line is indeed relevant, though not primarily for the feature intended by Masica. Double causatives do exist to the east of it. East of the 84th meridian, two areas can be set off: (A) a former contact zone between TB, AA, and DRAV, stretching from eastern Nepal to Orissa, and (B) the predominantly TB northeast. Only a few of the more than 100 languages of those tribal areas have been described, and hardly any of them has been considered in areal studies. The Non-IA languages from Nepal to Orissa (zone A) are characterized by a complex verbal morphology, which is not characteristic of the TB relatives farther north and east and may be due either to MU, which seems to have had a complex morphology at all times (Zide and Anderson, 2001), or to an unidentified third substratum (referred to, e.g., in Hook, 2001: 124; Witzel, 1999: 40). Different from the converbal structures typical of OV languages, much of the complex pattern of person and tense-aspect marking is retained in subordination in MU languages, in Kurukh, and in Kiranti languages of eastern Nepal. Sometimes only the finite or final marker is missing, as in Santali and Athpare. It seems that long contact between DRAV, MU, TB, and possibly other language groups have lead to an area little affected by the rest of South Asian language developments (Ebert, 1999: 392).

Subareas have been claimed for the Khondmals, for Jharkhand, for the Northwest Frontier, and others. This does not invalidate the hypothesis of a South Asian linguistic area, just as sharing the OV characteristics with Central Asia does not. We should not be surprised to find subareas within subareas without

clear boundaries. “Sprachbund situations are notoriously messy,” as Thomason and Kaufman (1988: 95) put it. But we should not look for “pan-Indic and not extra-Indic” features, as Emeneau suggested. Instead, research must concentrate on more detailed investigations of certain phenomena, such as compound verbs or converbs, which show uneven frequencies and idiosyncratic forms in subareas. The clustering of features in certain subareas is characteristic of linguistic areas, just as is their diffusion into a number of unrelated languages.

Bibliography

- Apte M L (1974). ‘Pidginization of a lingua franca: A linguistic analysis of Hindi-Urdu spoken in Bombay.’ In Southworth F C & Apte M L (eds.), 21–41.
- Bhaskararao P & Subbarao K V (eds.) (2001). *The yearbook of South Asian language and linguistics 2001. Tokyo symposium on South Asian languages. Contact, convergence and typology*. New Delhi: Sage Publications.
- Ebert K H (1999). ‘Kiranti non-finite forms – An areal perspective.’ In Yadava Y P & Glover W W (eds.) *Topics in Nepalese linguistics*. Kathmandu: Royal Nepalese Academy. 371–400.
- Emeneau M B (1956). ‘India as a linguistic area.’ *Language* 32(1), 3–16. [also 1980: 1–18].
- Emeneau M B (1980). *Language and linguistic area. Essays by Murray B. Emeneau*. Stanford, CA: Stanford University Press.
- Gumperz J J & Wilson R (1971). ‘Convergence and creolization: A case from the Indo-Aryan/Dravidian border.’ In Hymes D (ed.), 151–167.
- Hock H H (2001). ‘Typology vs. convergence: The issue of Dravidian/Indo-Aryan syntactic similarities revisited.’ In Bhaskararao P & Subbarao K V (eds.), 63–99.
- Hook P E (1987). ‘Linguistic areas: Getting at the grain of history.’ In Cardona G & Zide N (eds.) *Festschrift for Henry Hoernigswald*. Tübingen: Narr. 155–168.
- Hook P E (2001). ‘Where do compound verbs come from? (And where are they going?)’ In Bhaskararao P & Subbarao K V (eds.), 101–130.
- Hymes D (ed.) (1971). *Pidginization and creolization of languages*. Cambridge: Cambridge University Press.
- Kuiper F B J (1967). ‘The genesis of a linguistic area.’ *Indo-Iranian Journal* 10, 81–102. [Repr. In Southworth F C & Apte M L (eds.) (1974). 135–153.]
- Masica C P (1976). *Defining a linguistic area: South Asia*. Chicago: The University of Chicago Press.
- Masica C P (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Masica C P (2001). ‘The definition and significance of linguistic areas: Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area).’ In Bhaskararao P & Subbarao K V (eds.), 205–267.
- Ramanujan A K & Masica C (1969). ‘Toward a phonological typology of the Indian linguistic area.’ In Sebeok T E (ed.) *Current Trends in Linguistics, vol. 5: Linguistics in South Asia*. 543–577.
- Southworth F C (1971). ‘Detecting prior creolization. An analysis of the historical origins of Marathi.’ In Hymes D (ed.), 255–273.
- Southworth F C (1974). ‘Linguistic stratigraphy of North India.’ In Southworth F C & Apte M L (eds.), 201–223.
- Southworth F C & Apte M L (eds.) (1974). *Contact and convergence in South Asian languages. The Issue of International Journal of Dravidian Linguistics* 3(1).
- Steever S B (1993). *Analysis to synthesis. The development of complex verb morphology in the Dravidian languages*. Oxford: Oxford University Press.
- Thomason S G & Kaufman T (1988). *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.
- Trivedi G M (1990). ‘Echo formation.’ In Krishan S (ed.) *Linguistic traits across language boundaries*. Calcutta: Anthropological Survey of India. 51–82.
- Verma M K & Mohanan K P (eds.) (1990). *Experiencer subjects in South Asian languages*. Stanford: Stanford University.
- Witzel M (1999). ‘Substrate languages of Old Indo-Aryan (Rgvedic, Middle and Late Vedic).’ *Electronic Journal of Vedic Studies* 5(1), 1–67.
- Zide N H & Anderson G D S (2001). ‘The Proto-Munda verb system and some connections with Mon-Khmer.’ In Bhaskararao P & Subbarao K V (eds.), 517–540.

South Philippine Languages

S Brainard, Summer Institute of Linguistics,
Philippines, Manila, Philippines

© 2006 Elsevier Ltd. All rights reserved.

Introduction

South Philippine languages form one of three major language groups found in the Philippines, all of which

belong to the Western Malayo-Polynesian branch of the Austronesian family. The South Philippine language family includes the Subanon, the Danao, and the Manobo subgroups. The Subanon subgroup is spoken on the Zamboanga peninsula of Mindanao, the Danao subgroup is spoken in central Mindanao, and the Manobo subgroup is spoken in central and eastern Mindanao. Two other groups of languages spoken in the Philippines are not South Philippine languages but

Table 1 South Philippine, South Mindanao, and Sama languages^a

| <i>South Philippine</i> | <i>South Mindanao</i> | <i>Sama</i> |
|--|--|---|
| Subanon subgroup | Bilic | Sama |
| Eastern Subanon – Central Subanen, Northern Subanen, Lapuyan Subanon | Blaan – Koronadal Blaan, Sarangani Blaan | Sibuguey Sama, Northern Sama, Western Sama, Central Sama, Southern Sama |
| Kalibugan – Kolibugan Subanon, Western Subanon | Tboli | |
| Danao subgroup | Tiruray | Yakan |
| Maguindanao – Maguindanaon | Tiruray | Yakan |
| Maranao-Iranon – Ilanun, Maranao | Bagobo | Jama Mapun |
| Manobo subgroup | Giangan | Jama Mapun |
| North Manobo – Binukid, Higaonon, Kagayanen, Cinamiguin Manobo | | Abaknon |
| Central Manobo | | Abaknon |
| East Central Manobo – Agusan Manobo, Dibabawon Manobo, Rajah Kabunsuwan Manobo | | |
| South Central Manobo – Obo Manobo | | |
| Ata-Tigwa – Ata Manobo, Matigsalug Manobo | | |
| West Central Manobo – Ilianen Manobo, Western Bukidnon Manobo | | |
| South Manobo – Cotabato Manobo, Sarangani Manobo, Tagabawa Manobo | | |

^aBased on data from McFarland (1980).

are more likely related to languages of Indonesia and Malaysia. These languages, found in the southern Philippines, are the South Mindanao languages, spoken in southern Mindanao, and the Sama languages, spoken on the Zamboanga peninsula and the Sulu Archipelago (see **Table 1**).

Although considerable research has been devoted to identifying and grouping languages of the southern Philippines, less effort has been spent on describing the grammar of these languages. Most of the available descriptions, completed in the 1960s and 1970s, emphasized phonology, verb morphology, sentence structure, and discourse features. A number of important findings from cross-linguistic studies in the 1980s resulted in significant reanalyses of the basic verbal sentences of Philippine languages. The outcome is that the sentence type traditionally called the ‘goal-focus’ construction was confirmed to be the basic transitive sentence. Later studies have also suggested that the ‘actor-focus’ construction is not one but two distinct sentence types: an active intransitive sentence (in which the verb is semantically intransitive) and an antipassive construction (in which the verb is semantically transitive). The few descriptions of languages in the southern Philippines completed after the 1970s generally reflect these reanalyses and also provide more information about syntactic processes, giving attention to the function and behavior of constructions as well as their structure.

In the following discussion, grammatical relations are labeled as follows: the only grammatical relation of a single-argument sentence is S, the more

agent-like grammatical relation of a transitive sentence is A, and the less agent-like grammatical relation of a transitive sentence is O. Case markers are morphemes that formally distinguish between A and O in a transitive sentence. These markers form three patterns: nominative–accusative (henceforth ‘nominative’), ergative–absolutive (henceforth ‘ergative’), and tripartite. In the nominative pattern, S and A are marked the same, but O is marked differently; in the ergative (ERG) pattern, S and O are marked the same, but A is marked differently; in the tripartite pattern, S, A, and O are each marked differently. In a split ergative pattern, S, A, and O display ergative case marking in some sentences and nonergative case marking in others. If S, A, and O are all marked the same, the forms are said to be neutralized for case.

Phonology

South Philippine, South Mindanao, and Sama languages have relatively straightforward phonemic inventories. Vowel systems have four, five, six, or, more rarely, seven vowels. Consonants consist of voiced and voiceless stops (including glottal stop), a few fricatives (often /s/ and /h/), nasals, /l/ and a rhotic (which, depending on the language, is a flap, /ɾ/ or /ɽ/, or a trill, /r/), and the semivowels /w/ and /j/. Some of the Sama languages also have /ɖʒ/. Long vowels and geminate consonants are common. Word stress occurs on the penultimate or, less commonly, the ultimate syllable and may or may not be predictable, depending on the language.

In the Subanon and Manobo subgroups and in the Sama languages, intervocalic consonants (C), particularly /l/, tend to be deleted over time. This appears to be the source for two common phonological features noted in these languages: long vowels (V) and the syllable V(C). In a significant number of Philippine languages, an epenthetic glottal stop is inserted in syllable onsets (but not codas) when no other phonetic material is available; however, this strategy is apparently not available when an intervocalic consonant is lost in these languages. Thus, the loss of an intervocalic consonant opens a pathway for the resulting sequence of vowels to become long vowels or to coalesce into a new vowel: e.g., /o/ + /i/ → /e/ in Obo Manobo (Khor and Vander Molen, 1996: 31) and /a/ + /i/ → /æ/ and /a/ + /u/ → /ɔ:/ in Northern Subanen (Daguman and Sanicas-Daguman, 1997: 103). The loss of /l/ or some other intervocalic consonant also seems to be the source for word-medial V(C) syllables, also noted for Subanon and Manobo subgroups and Sama languages. Only Maguindanaon allows a word-initial V(C) syllable (the facts are not available for Maranao). Thus, syllable types for Maguindanaon, Obo Manobo, Northern Subanen, and Southern Sama are CV(C) and V(C).

On the other hand, the loss of a mid-central vowel in the South Mindanao subgroup has led to the development of word-initial CCV(C) syllables. For Tboli, the loss appears to have occurred in words beginnings with CV syllables. For Blaen, it occurred in words beginning with CVC syllables. This claim is based on the fact that in Tboli, a mid-central vowel can be inserted optionally after the first consonant of the root (e.g., *btang* ~ *betang* 'to fall') (Forsberg, 1992: 6), but in Blaen, the mid-central vowel is inserted optionally before the first consonant of the root (e.g., *bgang* ~ *ebgang* (adj.) 'broken') (Sally Winter, personal communication). Thus, syllable types for South Mindanao languages are CV(C) and CCV(C). Word-initial CCV(C) syllables also occur in Maranao but are limited to homorganic nasals followed by stops: e.g., /mb/, /nd/, /ŋk/, /ŋg/ (no occurrences of /mp/ or /nt/ are listed in McKaughan and Al-Macaraya (1996)).

Three other phonological features are also notable. One is a unique set of phonological alternations triggered by the syntactic category marker G- in Subanon languages. These alternations involve change in voicing, nasalization, point of articulation, spirantization, and deletion of G-, depending on the identity of the following consonant (Sanicas-Daguman, 1996: 63–64) (see later, Morphology). The second notable feature is neutralization of contrast between /a/ and certain other vowels in Manobo languages. Specifically, in several Manobo languages, /a/ contrasts with

all vowels only in the last two syllables of a word. It never occurs in any syllable to the left of the penultimate. When /a/ moves into one of these positions, it is replaced by one particular vowel: e.g., /a/ → /o/ in Obo Manobo (Khor and Vander Molen, 1996: 30), /a/ → /e/[ə] in Western Bukidnon Manobo (Elkins, 1963), and /a/ → /ɔ/ in Matigsalug Manobo (Elkins, 1984). The third feature is the tendency for high vowels in Maguindanaon to lose their moras under certain conditions, such that high vowels surface as palatalization and labialization on preceding consonants (if they are the first vowel in the sequence) or as glides of diphthongs (if they are the second vowel in the sequence) (Lee, 1964; Skoropinski, 2004).

Morphology

Morphology tends to be more complex in South Philippine languages than in South Mindanao and Sama languages. Taking verbal morphology as an example, South Philippine languages display a fairly extensive range of verbal affixes, some of which undergo complex phonological alternations. These affixes function as portmanteau morphemes, signaling a variety of syntactic and semantic information. The most common information is transitivity (intransitive vs. transitive), dynamism (dynamic vs. stative), and the semantic role of S or O. Aspect, mood, and, less commonly, tense are marked on the verb, but for at least one South Philippine language (Obo Manobo), a mood contrast (realis vs. irrealis) is signaled by clause-level clitics, as well as by verb affixes. Other types of information that may be marked on the verb are abilitative, habitual, reciprocal, distributive, and multiple participants. South Mindanao and Sama languages typically take fewer verbal affixes, and at least some tense, aspect, and mood contrasts are indicated by words or clitics, rather than by verbal affixes. On the other hand, Sama languages gain in complexity through the ubiquitous presence of the affix *pa-*, which attaches to many verb stems and performs several functions. Most commonly, *pa-* creates new words and adds arguments to the sentence (e.g., an agent to a basic verbal sentence or a causer to a causative construction).

Three other morphemes are also of interest. The first is the Subanon syntactic category marker G-, undoubtedly the single most interesting morpho-phonological feature of this subgroup. The marker G- attaches to nouns and to all lexical constituents of noun phrases (NPs), marking the constructions as nominals. Evidence suggests that G- is the final consonant of an old case marker that over time became phonologically attached to the first consonant of the following nominal; it has subsequently

been reanalyzed as a syntactic category marker, i.e., a nominal marker in the literal sense.

The second morpheme of note is the verbal affix *-an*. This affix occurs in most Philippine languages and functions as a valence increaser, i.e., it occurs on a verb when an oblique NP, usually a location, a recipient, or a beneficiary, is promoted to the O argument (direct object) (see later, Syntactic Processes). Although *-an* performs this function in certain Sama languages (Southern Sama and Balangingi Sama), in these languages it also has a second function – that of verb classifier. As a verb classifier, *-an* occurs on some but not all verbs when O is a patient (PAT). (For most semantically transitive verbs, a patient is the unmarked choice for O; consequently, *-an* cannot be functioning as a valence increaser when it cross-references an O patient.) Since *-an* occurs on some, but not all, verbs when O is a patient, it divides verbs into two classes, those that require *-an* when O is a patient and those that do not. This function of *-an* seems to be unique to the Sama subgroup.

The third morpheme is the Yakan clitic *-in*. This clitic occurs on NPs and nominalized sentences and has two functions. First, it signals that a nominal is definite (DEF). In Yakan transitive (TRANS) sentences, O must be definite but not A. Consider Example (1) (Brainard and Behrens, 2002: 42):

- (1) tinnennun we' dende bunga-samahin
 -in-tennun we' dende bunga-sama-in
 TRANS-weave ERG woman bunga-sama.type-DEF
 'a woman wove the *bunga-sama* type of weaving.'

Second, *-in* marks S of a single-argument sentence (i.e., term (TRM)), whether or not S is definite (Example (2)) (Brainard and Behrens, 2002: 52):

- (2) lakkes kura'-in
 fast horse-TRM
 'a/the horse is fast.'

A phonologically identical morpheme that appears to have a similar function also occurs in Balangingi Sama (Gault, 1999: 18).

Morphosyntax of Basic Sentence Types

South Philippine, South Mindanao, and Sama languages display typical Philippine-type sentence structure: sentences are verb-initial and main verbs usually take an affix that cross-references one, and only one, NP in the sentence (i.e., S in a single-argument sentence and O in a transitive sentence). Sentences that express identity, attribute, possession, location, and existence are usually verbless. Sentences that express states may pattern like verbless sentences (in which case the state is coded as an adjective), or like verbal

sentences (in which case the state is coded as a stative verb). Sentences that express actions are verbal sentences and are grouped into four types: active intransitive, transitive, antipassive, and passive. The active intransitive sentence has a semantically intransitive verb; all the other sentence types have semantically transitive verbs (or verbs that pattern like semantically transitive verbs).

In some Philippine languages, transitive sentences have two word orders: VAO and VOA. Traditionally, this pattern has been explained in terms of phonology (i.e., the phonologically shorter argument precedes the phonologically longer one) or in terms of morphology or topicality (i.e., a pronoun precedes a full NP); however, neither explanation has accounted for all the facts. Brainard and Vander Molen (2003) suggested that the VOA sentence is a word-order inverse (a voice construction first proposed by Givón (1994)). Selection of the VOA inverse construction over the VAO active construction is determined either by a person hierarchy (if only first and second persons are involved) or a topicality hierarchy (if only third persons are involved), or a combination of both hierarchies (if first, second, and third persons are all involved). If full NPs as well as pronouns are involved in the selection, then the hierarchy looks like that in **Figure 1**. In general, if A outranks O on the hierarchy, the VAO active construction is selected, but if O outranks A, the VOA inverse construction is selected.

Antipassives and passives are detransitivized constructions, i.e., constructions in which one grammatical relation of a transitive sentence has been demoted to oblique or deleted. In Philippine antipassives, O of the transitive counterpart is demoted or deleted. Although the demoted NP is often indefinite, it may be definite. Following demotion or deletion of O, A becomes S. In a Philippine passive, A of the transitive counterpart is obligatorily deleted, and following deletion, O becomes S.

Two types of passives occur in Philippine languages: a morphological passive and a nonmorphological passive. In the morphological passive, the verb takes stative affixes, but in the nonmorphological passive, the verb takes the same affixes that occur on it in a transitive sentence. Thus, the only difference between a nonmorphological passive and a transitive sentence is the obligatory absence of A in the passive. As it happens, some Philippine languages have both types of passives. The factors determining the selection of

First person > Second person > Third person > Pronouns > Full noun phrases

Figure 1 Person-topicality hierarchy governing VAO and VOA selection.

one passive over the other appear to be language specific and have not yet been fully investigated.

With the identification of the goal-focus construction as the basic transitive sentence, case-marking patterns have undergone reexamination. In South Philippine and Sama languages, case marking displays either a consistently ergative pattern or a split ergative pattern (the precise details of the split ergative patterns vary from language to language). In South Mindanao languages, nominal markers do not function as case markers, although pronouns are marked for case.

At this point, it may be useful to compare actual data from representative languages. When discussing nominal markers, only the marking of S, A, and O will be considered.

Maguindanaon

Common nouns and personal names are marked for case and display an ergative pattern (see **Table 2**) (it is unclear if the VOA inverse is possible when A and O are both full NPs). Pronouns are also marked for case (see **Table 3**). In a VAO active construction, second-person pronouns have a tripartite pattern, but third-person pronouns have an ergative pattern. In a VOA inverse construction, second-person pronouns have an ergative pattern (for all other persons, either

A or O does not occur in the construction). Maguindanaon has five types of verbal sentences: active intransitive, VAO active construction, VOA inverse construction, antipassive, and passive. Example (3) is an active intransitive sentence (Bruce Skoropinski, personal communication):

- (3) lemu aku saguna
 leave 1SG now
 'I will leave now.'

Selection of a VAO active construction (Example (4); Bruce Skoropinski, personal communication) and a VOA inverse construction (Example (5); Fleischman, 1986: 30) is governed by a person-topicality hierarchy identical to that in **Figure 1** (in the following examples, COMP means 'completed aspect'):

- (4) in-umbal-an ku seka sa liplanu
 COMP-make-BEN 1SG 2SG OBL airplane
 'I made you an airplane.'
- (5) in-enggat aku nengka kanu walay nengka
 COMP-invite 1SG 2SG OBL house GEN.2SG
 'you invited me to your house.'

Example (6) (Fleischman, 1986: 30) is the antipassive. (Example (7) is its transitive counterpart.)

- (6) min-umbal aku sa liplanu sa leka
 COMP-make 1SG OBL airplane OBL OBL.2SG
 'I made an airplane for you.'
- (7) in-umbal ku su liplanu sa leka
 COMP-make 1SG ABS airplane OBL OBL.2SG
 'I made the airplane for you.'

The Maguindanaon passive is a nonmorphological passive. Compare the passive in Example (8) (Bruce Skoropinski, personal communication) with its transitive counterpart in Example (6):

- (8) in-umbal su liplanu sa leka
 COMP-make ABS airplane OBL OBL.2SG
 'the airplane was made for you'

Obo Manobo

Obo Manobo has two types of transitive sentences, a VAO active construction and a VOA inverse construction. Case marking of common nouns and personal names in both constructions is identical and displays a consistently ergative pattern (see **Table 4**). Pronouns are also case marked (see **Table 5**). In a VAO active construction, first- and second-person pronouns have a tripartite pattern and third-person pronouns have an ergative pattern. On the other hand, in a VOA inverse construction, first-person plural exclusive pronouns and second- and third-person pronouns have an ergative pattern. (The pronouns *nikoddi* '1SG' and *niketa* '1PL.INCL' have recently come to notice and also appear

Table 2 Maguindanaon case markers

| Noun type | Marker | | |
|----------------|--------|----|----|
| | S | A | O |
| Common nouns | | | |
| Definite | su | nu | su |
| Indefinite | i | na | i |
| Personal names | si | ni | si |

Table 3 Maguindanaon pronouns

| Person/ number | VS sentence, S | VAO sentence | | VOA sentence | |
|-------------------|-------------------|--------------|---------|--------------|--------|
| | | A | O | O | A |
| Singular | | | | | |
| 1 | aku | ku | – | aku | – |
| 2 | ka | nengka | seka | ka | nengka |
| 3 | sekanin | nin | sekanin | – | nin |
| Plural | | | | | |
| 1INCL | tanu | tanu | – | tanu | – |
| 1DU | ta | ta | – | ta | – |
| 1EXCL | kami | nami | – | kami | – |
| 2 | kanu | nu | sekanu | kanu | nu |
| 3 | silan | nilan | silan | – | nilan |

to be possible for A in VOA constructions, although this needs to be confirmed.) Obo Manobo has six types of verbal sentences: intransitive, VAO active construction, VOA inverse construction, antipassive, morphological passive, and nonmorphological passive. Example (9) (Edna Vander Molen, personal communication) is an active intransitive sentence:

- (9) od usok ka diyon to baoy
IRR enter 2SG there OBL house
'you will enter into the house'

Examples (10) and (11) (Vera Khor, personal communication) show, respectively, a VAO active construction and a VOA inverse construction:

- (10) od suntuk-on din sikkow
IRR hit-PAT 3SG 2SG
'he will hit you'
- (11) od suntuk-on ka nikandin
IRR hit-PAT 2SG 3SG
'he will hit you'

Selection of the active construction and the inverse construction is controlled by a person-topicality hierarchy identical to that in Figure 1. Obo Manobo is notable in that both constructions are possible for most person combinations. Word-order inverses have also been noted for Agusan Manobo, Matigsalug Manobo, Sarangani Manobo, Tagabawa Manobo,

and Western Bukidnon Manobo. Example (12) is the antipassive. (Example (13) is its transitive counterpart (Ena Vander Molen, personal communication):

- (12) od tampod iddos anak to tali
IRR cut ABS child OBL rope
'the child will cut a rope'
- (13) od tompoddon to anak iddos tali
od tampod-on to anak iddos tali
IRR cut-PAT ERG child ABS rope
'the child will cut the rope'

Examples (14) and (15) (Ena Vander Molen, personal communication) are, respectively, a morphological passive and a nonmorphological passive (Example (13) is the transitive counterpart):

- (14) od ko-tampod iddos tali
IRR PASS-cut ABS rope
'the rope will be cut'
- (15) od tompoddon iddos tali
od tampod-on iddos tali
IRR cut-PAT ABS rope
'the rope will be cut'

Tboli

Tboli common nouns and personal names are not marked for case, but pronouns are and display a split case-marking system (see Table 6) (some pronouns have allomorphs, not all of which are listed in Table 6). The first split occurs between singular and plural forms. Singular forms have a tripartite pattern. The second split occurs between plural forms: all plural forms except first-person inclusive have a nominative pattern, and first-person inclusive forms are neutralized for case. Table 6 shows the distribution of pronouns in affirmative sentences. A notable feature of Tboli is that the negation of a sentence triggers a change in pronoun sets for S and O. This change also alters the case-marking pattern slightly (see Table 7). In negated sentences, singular first and second persons still display a tripartite pattern, but singular

Table 4 Obo Manobo case markers

| Noun type | Marker | | |
|----------------|-----------------|------------|-----------------|
| | S | A | O |
| Common nouns | | | |
| Definite | idda (so) | (tadda) to | idda (so) |
| General | ko/do | to | ko/do |
| Specific | ko (so)/do (so) | to | ko (so)/do (so) |
| Personal names | | | |
| Singular | si | ni | si |
| Plural | onsi | onni | onsi |

Table 5 Obo Manobo pronouns

| <i>Person/number</i> | <i>VS sentence, S</i> | <i>VAO sentence</i> | | <i>VOA sentence</i> | |
|----------------------|-----------------------|---------------------|----------|---------------------|----------|
| | | <i>A</i> | <i>O</i> | <i>O</i> | <i>A</i> |
| Singular | | | | | |
| 1 | a | ku | siyak | a | – |
| 2 | ka | du/ru | sikkow | ka | nikkow |
| 3 | sikandin | din/rin | sikandin | sikandin | nikandin |
| Plural | | | | | |
| 1INCL | ki | ta | siketa | ki | – |
| 1EXCL | koy | doy/roy | sikami | koy | nikami |
| 2 | kow | dow/row | sikiyu | kow | nikiyu |
| 3 | sikandan | dan/ran | sikandan | sikandan | nikandan |

Table 6 Distribution of pronoun sets in affirmative Tboli sentences^a

| Pronoun | Person/number | | |
|----------|---------------|-------|-------|
| | S | A | O |
| Singular | | | |
| 1 | -e | -u | ou/o |
| 2 | -i | -em | uu/u |
| 3 | ø | -en | du |
| Plural | | | |
| 1INCL | tekuy | tekuy | tekuy |
| 1DU | te | te | tu |
| 1EXCL | me | me | mi |
| 2 | ye | ye | yu |
| 3 | le | le | lu |

^aBased on data from Forsberg (1992: 22), with permission.**Table 7** Distribution of pronoun sets in negated Tboli sentences

| Pronoun | Person/number | | |
|----------|---------------|-------|--------|
| | S | A | O |
| Singular | | | |
| 1 | -e | -u | dou/do |
| 2 | -i | -em | kóm |
| 3 | -en | -en | du |
| Plural | | | |
| 1INCL | tekuy | tekuy | tekuy |
| 1DU | te | te | kut |
| 1EXCL | me | me | kum |
| 2 | ye | ye | kuy |
| 3 | le | le | kul |

third persons now display a nominative pattern. All other persons except first-person plural inclusive continue to display a nominative pattern; first-person plural inclusive continues to be neutralized for case (the alternation of pronouns in affirmative and negated sentences does not occur in Blaan). Examples (16)–(19) illustrate the changes in pronouns. When a single-argument sentence is negated, S changes only when it is a singular third person (Examples (16) and (19) from Lillian Underwood (personal communication); Examples (17) and (18) from Forsberg (1992: 101, 102)):

(16) mung-e
go-1SG
'I'm going along'

(17) là mung-e
not go-1SG
'I'm not going along'

(18) mung
go
'he is going along'

(19) là mung-en
not go-3SG
'he is not going along'

When a transitive sentence is negated, O changes when it is any person except third-person singular and first-person plural inclusive (Examples (20) and (21); Porter, 1977: 114, 115):

(20) nwit Kasi ou elem bulul
TRANS.take Kasi 1SG to mountains
'Kasi took me to the mountains'

(21) là nwit Kasi dou elem bulul
not TRANS.take Kasi 1SG to mountains
'Kasi didn't take me to the mountains'

As might be expected, word order is relatively rigid and a primary means of distinguishing between A and O in transitive sentences; however, when S, A, or O is an expanded full NP, the NP moves to the end of the sentence, and a coreferential pronoun is left in the normal sentence position (Example (22); Forsberg, 1992: 57) (in the following example, PREP means 'preposition'):

(22) kól le_i bélé me
arrive 3PL PREP 1PL.EXCL
[kem tau dmadu]_i
PL person INTRANS.plow
'the men who are to plow have arrived to us'

If both A and O are expanded NPs, both NPs move to the end of the sentence, with A coming last. Coreferential pronouns are left for A and O in their normal positions (Example (23); Porter, 1977: 99) (in the following example, SPEC means 'specific'):

(23) eted le_i lu_j [yó kem
deliver 3PL 3PL SPEC PL
ngà lemnek]_j [yó kem
child small SPEC PL
tau lem wót gu leged]_i
person INTRANS.come from upstream
'the people from upstream delivered the small children'

If only one expanded NP is present at the end of a transitive sentence, it always refers to O (Example (24); Porter, 1977: 98):

(24) eted le lu_i [yó kem ngà lemnek]_i
deliver 3PL 3PL SPEC PL child small
'they delivered the little children'

Expanded NPs in Blaan do not change sentence position.

Southern Sama

Southern Sama common nouns and personal names display a consistently ergative case-marking pattern.

Table 8 Southern Sama pronouns

| Pronoun | Person/number | | |
|----------|---------------|------|--------|
| | S | A | O |
| Singular | | | |
| 1 | akú | ku | akú |
| 2 | kow | nu | kow |
| 3 | iyá | na | iyá |
| Plural | | | |
| 1INCL | kitabí | tabí | kitabí |
| 1DU | kitá | ta | kitá |
| 1EXCL | kamí | kâmi | kamí |
| 2 | kam | bi | kam |
| 3 | sigá | sigá | sigá |

For common nouns and personal names, S and O have no case marker, but A is obligatorily marked by *heb*. For pronouns, all persons except plural third persons display an ergative pattern (see Table 8). Plural third-person pronouns are neutralized for case. Note that phonological contrast is minimal for first-person plural exclusive: S, A, and O are identical except for word stress (represented in Table 8 by an acute accent). When A is a pronoun in a transitive sentence of type 1 (see Examples (26) and (27)), it is also obligatorily marked by *heb*. Southern Sama has five types of verbal sentences: active intransitive, transitive type 1, transitive type 2, antipassive, and passive. Example (25) is an active intransitive sentence (Trick, 1997: 126):

- (25) pasód anak-anak ni lumah
 enter child OBL house
 'the child will enter into the house'

Transitive sentences are of two types. Transitive sentence type 1 is more morphologically complex, compared to type 2, because the verb must occur with the affix *ni-* (or its allomorph *-in-*), A is either a full NP or a pronoun and must be preceded by the ergative marker *heb*, and word order may be VAO or VOA, with VOA being the more common order (transitive type 1, Examples (26) and (27), VOA and VAO order, respectively; Trick, 1997: 128):

- (26) sinampak eroh heb anak-anak
 sampak-in- eroh heb anak-anak
 slap-TRANS dog ERG child
 'the child will slap the dog'
- (27) sinampak heb anak-anak eroh
 sampak-in- heb anak-anak eroh
 slap-TRANS ERG child dog
 'the child will slap the dog'

Transitive sentence type 2 is morphologically simpler: the verb never occurs with *ni-*, A must be a pronoun and is never preceded by *heb*, and word order is

obligatorily VAO (Example (28); Doug Trick, personal communication) (similar pairs of transitive sentences have also been noted for Balangingi Sama, Pangutaran Sama, and Yakan):

- (28) sampak-ku eroh
 slap-ERG.1SG eroh
 'I will slap the dog'

Example (29) (Trick, 1997: 132) is the antipassive. (Example (30) (Trick, 1997: 132): is its transitive counterpart.)

- (29) ngan-dugsuh aku sowa
 AGT-stab ABS.1SG snake
 'I will stab a/the snake'
- (30) ni-dugsu-an sowa heb-ku
 TRANS-stab-PAT snake ERG-ERG.1SG
 'I will stab the snake'

Passive sentences in Southern Sama are nonmorphological passives (Example (31); Trick, 1997: 133); compare Example (31) with its transitive counterpart, Example (29):

- (31) ni-dugsu-an sowa
 TRANS-stab-PAT snake
 'the snake will be stabbed'

Syntactic Processes

Although syntactic processes have been investigated in a few languages, e.g., Sama, Yakan, and Northern Subanen, this is an area of Philippine linguistics that still needs more research. What has been noted to date is that South Philippine, South Mindanao, and Sama languages, like all Philippine languages, allow an oblique NP to be promoted to O (i.e., direct object), although languages vary as to which semantic roles may undergo promotion. The promoted NP is always cross-referenced by an affix on the verb. A variation of this process occurs in some Sama languages. In cleft constructions in Philippine languages, S and O are the only arguments eligible to be the head of the construction, in which case they are usually cross-referenced by a verbal affix. For certain verbs, however, the verbal affix may cross-reference an oblique NP. For these NPs, morphological and syntactic evidence shows that the cross-referenced NP has changed its relation to the verb, but has not become a grammatical relation (i.e., O). (This process has been noted for Southern Sama and Yakan. In Southern Sama, a similar process also occurs in antipassive constructions.) For those languages in which other syntactic processes have been described (e.g., relativization, clefting, raising, coreferential deletion, and control of second-position clitics), the following

preliminary generalization can be made: in South Philippine languages, control of syntactic processes seems to be more or less evenly distributed between A and O in transitive sentences, but in Sama languages, control for nearly all of these processes (including second-position clitics) is governed exclusively by O in transitive sentences, making the Sama languages highly syntactically ergative languages. As for all Philippine languages, S is always the syntactic control in single-argument sentences.

Bibliography

- Abrams N (1961). 'Word base classes in Bilaan.' *Lingua* 10(4), 391–402.
- Abrams N (1970). 'Bilaan morphology.' *Papers in Philippine Linguistics* No. 3. *Pacific Linguistics, Series A* No. 24, 1–62.
- Brainard S & Behrens D (2002). *A grammar of Yakan. Special monograph issue no. 40* (vol. 1). Manila: Linguistic Society of the Philippines.
- Elkins R (1963). 'Partial loss of contrast between *a* and *e* in Western Bukidnon Manobo.' *Lingua* 12(2), 205–210.

- Forsberg V M (1992). 'A pedagogical grammar of Tboli.' *Studies in Philippine Linguistics* 9(1), 1–110.
- Gault J M (1999). *An ergative description of Sama Bangingi*. *Special monograph issue no. 46*. Manila: Linguistic Society of the Philippines.
- Kerr H (1988). 'Cotabato Manobo grammar.' *Studies in Philippine Linguistics* 7(1), 1–123.
- McKaughan H P & Al-Macaraya B (1996). *A Maranao dictionary* (rev. edn.). Manila: De La Salle University Press and Summer Institute of Linguistics.
- Sanicas Daguman J (2004). *A grammar of Northern Subanen*. Ph.D. thesis (unpubl.), La Trobe University, Bundoora, VI, Australia.
- Schlegel S A (1971). *Tiruray-English lexicon*. *University of California publications in linguistics*, vol. 67. Berkeley/Los Angeles/London: University of California Press.
- Sullivan R E (compiler) (1986). *A Maguindanaon dictionary*. Cotabato City, Philippines: Notre Dame University.
- Trick D (1997). 'Equi-NP deletion in Sama Southern.' *Philippine Journal of Linguistics* 28, 125–144.
- Walton C (1986). *Sama verbal semantics: classification, derivation and inflection*. Manila: Linguistic Society of the Philippines.

Southeast Asia as a Linguistic Area

W Bisang, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Mainland Southeast Asia – the Area, Its Languages and Language Families, Its History

Mainland Southeast Asia geographically covers the area of Vietnam, Laos, Cambodia, Thailand, Myanmar, peninsular Malaysia, and southern and southwestern China. This area is characterized by at least two millennia of lively exchange and interaction among speakers of languages that belong to no fewer than five families: Sino-Tibetan, Mon-Khmer (a subfamily of Austroasiatic), Tai (the core group of the Tai-Kadai languages), Hmong-Mien (also called Miao-Yao) and Chamic (Malayo-Polynesian subfamily of Austronesian). The families forming the core of mainland Southeast Asia as a zone of contact-induced convergence are Mon-Khmer, Tai, Hmong-Mien and Sinitic. The Hmong-Mien languages are divided into the Hmong (Miao) and the Mien (Yao) subfamilies. They are spoken in small areas of southern China and in northern Vietnam, Laos, and Thailand. The architecture of the other

families is presented in Tables 1–3: Table 1 is on Mon-Khmer, Table 2 on Tai, and Table 3 on Sinitic.

The present linguistic situation in mainland Southeast Asia is the result of extensive migrations, the rise and fall of many kingdoms, and innumerable contact situations (on the historical facts, see Wyatt, 1982). In the first millennium A.D., the inhabitants of this area had contacts with China and India. Vietnam, in the east, was governed by China between 179 B.C. and 938 A.D., while the west and the south were influenced by Theravada Buddhism through the mediation of the Mon (cf. the large corpus of Sanskrit and Pali words in modern Thai and Khmer). The geography of mainland Southeast Asia and its large rivers in particular directed migration from southern China toward Thailand, Laos, and Vietnam. Apart from Chinese and Indian influence, the first millennium A.D. is characterized by the steady emergence of greater political structures. At its end, we find the state of Vietnam, the kingdom of Champa (on the coast of central Vietnam; Austronesian: Chamic), the Khmer empire of Angkor, the kingdoms of central and northern Thailand, and the Burmese kingdoms of Mon and Pyu. At about the same time, numerous speakers of Tai languages migrated from inland southern China (Guizhou, Guangxi) to the south

Table 1 Subgrouping of Mon-Khmer (MK) languages (according to Diffloth & Zide, 1992)

| | | |
|--|---|---|
| Northern | Khmuic (N Laos/N Thailand) Palaungic (N Laos/N Thailand, E Burma, SW Yunnan) Khasian (NE India) | Khmu, Mal-Phrai, Mlabri Eastern: Riang dialects, Danau Western: Waic, Angkuic, Lametic Khasi |
| Eastern | Khmeric Bahnaric (35 languages in Central and S Vietnam, S Laos, E Cambodia) Katuic (Central Vietnam/Laos, NE Thailand, N Cambodia) Pearic (Central Cambodia; affiliation uncertain) | Khmer South: Srê, Mnong, Stieng, Chrau Central: Bahnar, etc. West: Brao (Lave), Nya-heuny (Nyaheun), etc. North: Rengao, Sedang, etc. West: Kuy, Bru, Sô, etc. East: Katu, Pacoh, Ngeq, etc. Samrê, Pear, Sa-och (Sa'och), Chong |
| Viet-Muong (is probably a branch of Eastern MK) | | Vietnamese, Muong, etc. |
| South | Monic Aslian (interior Malaysia, 16 languages) Nicobarese (Nicobar Islands, may be another direct branch of MK) | Mon (Myanmar, Thailand), Nyahkur (E Central Thailand) Senoic: Semai, Temiar; North: Kintaq, Jahai (Jehai), Batek; South: Mah Meri (Besisi), Semelai; Jah Hut Four subgroups |

Table 2 Subgrouping of Tai languages (according to Li, 1977; for more information, see Edmondson and Solnit, 1997)

| | |
|--------------|--|
| Southwestern | Ahom (Assam/India; extinct) Central Thai (= Siamese) East Central: Black Tai (Tai Dam) (N Vietnam), Red Tai (Tai Daeng) (North Central Vietnam), Phu Tai (Phuan) (Laos/Thailand) Khamti (NW Myanmar; Assam) Lanna (N Thailand): Mueang, Northern Thai, Yuan Lao (Laotian, including Isan in Thailand) Lue (Lü) (called Dai in China; situated in Yunnan) Shan (SE Myanmar, Thailand) Southern Thai White Tai (Tai Dón) (N Vietnam) etc. |
| Central Tai | Nung (on both sides of the Chinese-Vietnamese border), Southern Zhuang (China, Zhuang Autonomous Region), Tho (= Tay and Caolan; NE Vietnam, S China) |
| Northern Tai | Bouyei (Buyi), Saek (Central Laos near Vietnamese border, NE Thailand), Northern Zhuang (across S China) |

and changed the balance in the north towards the Tai population, which simultaneously became an important reservoir of manpower and a potentially dangerous rival for the adjacent kingdoms. Cambodia moved its center of gravity from Angkor further south to Phnom Penh, and the newly developed Thai kingdoms of Sukhothai (?1240–1438) and Ayudhya (1351–1767) were characterized by intensive contact and presumably by a considerable proportion of bilinguals. As a consequence, there is a high degree of structural similarity and some lexical similarity between the two languages. Finally, the

Table 3 List of Sinitic languages/dialects (Chappell, 2001: 6)

| |
|--|
| Northern Chinese (N China, W China, par of central China, Sichuan basin, Guizhou and Yunnan provinces) Xiāng (Xiang Chinese) (Hunan province) Gàn (Gan Chinese) (Jiangxi province) Wú (Wu Chinese) (coastal area of lower Yangze River in the provinces of Jiangsu, Zhejiang and Anhui) Mǐn (Min Nan Chinese) (Southern coastal province of Fujian and the island of Taiwan, Leizhou peninsula plus Hainan island) Kèjǐā or Hakka (Hakka Chinese) (scattered throughout SE China in small communities in the Yuè and Mǐn areas) Yuè (Yue Chinese) (Guangdong and Guangxi provinces; Cantonese) |
| Recently identified dialect groups: |
| Jīn (Jinyu Chinese) (Shanxi province and Inner Mongolia) Píng huà (Guangxi) Huī (Huizhou Chinese) (in parts of Anhui, Jiangxi and Zhejiang provinces) |

migration of a considerable number of Hmong-Mien from southern and southwestern China to Laos, Vietnam, and Thailand started in the middle of the 19th century.

Given the long-lasting and very complex patterns of interaction among speakers of a large number of languages from different families, structural convergence comes as no surprise. Studies dealing with mainland Southeast Asia from an areal perspective are Huffman (1973), Clark (1978), Capell (1979), Clark (1989), Matisoff (1991), Bisang (1992), Bisang (1996), and Enfield (2003). Huffman (1986) is an excellent bibliography on the languages and linguistics of this area.

General Properties of the Languages of Mainland Southeast Asia – the Relevance of Pragmatics

Mainland Southeast Asian languages are characterized by a high degree of indeterminateness, which as a consequence endorses the relevance of pragmatic inferencing and produces a special type of pragmatics-oriented grammaticalization (see ‘Indeterminateness and the Role of Pragmatics’ below). The pragmatics-oriented character of grammaticalization may be one reason for a syllable-based morphology (see ‘Syllabic Morphology’ below). Another consequence of this type of grammaticalization may be the comparatively weak correlation between the lexicon and individual lexical items (see ‘Versatility’ below).

The above pragmatics-based properties will be discussed in this section. Two additional general properties will be treated in the subsection ‘Directional Verbs, Coverbs and TAM Markers, and Syntactic Patterns’ with the necessary language-specific details. The properties are the existence of rigid syntactic patterns with fixed functionally determined positions and the functional motivations of these patterns.

Indeterminateness and the Role of Pragmatics

East and mainland Southeast Asian languages are well known for their indeterminateness, i.e., their lack of obligatory categories (Bisang, 1992, 2001, see also context dependency in Enfield, 2003: 55). One famous instance is the lack of obligatory arguments. In the following example from Modern Standard Chinese (Mandarin Chinese), the agent argument *wǒ* ‘I’ and the patient argument *tā* ‘he’ are no longer mentioned in the second clause with the predicate *jiàn* ‘see’ because they are already known from the previous context.

- (1) *wǒ₁ bú jiàn tā₂ yǐ shì sān*
I NEG see he already be 30
shíduō nián; jīntiān Ø₁ jiàn Ø₂ le
more year today see PF
 ‘I haven’t seen him for more than 30 years. Today [I]
 saw [him]’
 (from Lu Xun, *Kuangren riji* [*Diary of a Madman*],
 second sentence).

Dropping arguments (prodrop) is not the only instance of indeterminateness. There are also a large number of grammatical categories which are optional (cf. Table 4), i.e., the speaker is not committed to select a particular subcategory (e.g., past, present, or future) from a particular obligatory category (e.g., tense).

Indeterminateness implies that grammatical categories which are expressed obligatorily in other languages must often be inferred from the context in mainland Southeast Asian languages. If these

Table 4 Some nonobligatory categories in mainland SE Asian languages

| Verb | Noun |
|--|--|
| Person/Number | Number |
| Tense/Aspect/Modality (TAM) | Noun class |
| Transitivity (transitive vs. intransitive) | Reference (definite, specific, indefinite) |
| Diathesis | Relationality (possession) |
| Causativity | Case |

categories are expressed, however, they are very often expressed by lexical items which occur in a special syntactic position of a construction where they get reanalyzed as grammatical markers. This type of grammaticalization differs from grammaticalization as described in the literature by dint of the vast functional range of many markers (see ‘Classifiers’ and “The Verb ‘Come to Have’” below) and the lack of a form–meaning correspondence (see ‘Directional Verbs, Coverbs and TAM Markers, and Syntactic Patterns’). The mainland Southeast Asian languages show that a high degree of abstraction (semantic generality, cf. Bybee, 1985) does not automatically lead to morphological reduction (see ‘Syllabic Morphology’ below); in other words, the semantic integrity of a linguistic sign is not fully reflected in its phonological integrity (Lehmann, 1995). The high functional range of individual markers is sometimes observed within an individual language, sometimes across languages. In the latter case, individual languages select certain domains out of the whole inferential potential of a marker common to a wider contact zone (see ‘Directional Verbs, Coverbs and TAM Markers, and Syntactic Patterns’ on verbs with the meaning ‘finish’ and the end of “The Verb ‘Come to Have’”).

The high relevance of pragmatics led to a more general discussion of the relevance of syntax in mainland Southeast Asian languages. Diller (1988) talked about “pragmatically organised syntax” in the context of Thai and other languages. Huang (1994) argued, against Huang (1984), that the interaction of syntax and pragmatics is subject to typological variance:

There seems to exist a class of language (such as Chinese, Japanese, and Korean) where pragmatics appears to play a central role which in familiar European languages (such as English, French, and German) is alleged to be played by grammar. In these ‘pragmatic’ languages ... (Huang, 1994: xiv)

Syllabic Morphology

Mainland Southeast Asian languages are characterized by or drift towards a morphology whose smallest meaningful element is the syllable. This definition

encompasses the well-known monosyllabism of languages such as Chinese (Mandarin Chinese) or Vietnamese (each syllable has its own meaning) but it also covers such cases as Thai or Khmer, which easily accept strings of semantically unanalyzable syllables, as in the elegant word for ‘restaurant’ in Thai (*pháttaakhaan*) or in Khmer (*phò:čəni:əttə:n*).

This does not mean that subsyllabic morphology does not exist in mainland Southeast Asia, but the integration into that area seems to engender a drift towards syllabic morphology. This can be illustrated by Mon-Khmer. Vietnamese, which was under the strong influence of monosyllabic Chinese for more than a millennium (see above), has completely lost its subsyllabic morphology, while Khmer, which had weaker contacts with China and was even able to transfer a lot of its vocabulary to Thai, probably has the richest morphology within the Mon-Khmer family (on Khmer morphology see Jenner and Pou, 1980–1981; Haiman, 1998). In spite of this, Khmer morphology is basically a lexical phenomenon, i.e., the affixes are not used productively. The productive strategies are all based on products of grammaticalization (as described in the subsections ‘Classifiers,’ ‘Directional Verbs, Coverbs and TAM Markers,’ and ‘Syntactic Patterns,’ and “The Verb ‘Come to Have’”). In addition, Khmer subsyllabic morphology is characterized by the following two properties: (1) a large number of Khmer affixes lack functional consistency, i.e., the same marker can express different functions depending on its base (the prefix *prə-* marks causativity/factivity, change of word class and reciprocity); and (2) the same function can be expressed by different affixes (e.g., derivation of nouns from verbs belongs to the functional range of the following affixes: *k-*, *s-*, *m-*, *N-*, *bvN-*, *kvN-*, *svN-*, *-b-*, *-m-*, *-n-*, *-vmn-/vN-*) (Bisang, 2001: 195–200).

Versatility

The term ‘versatility’ refers to the fact that the occurrence of a given linguistic item is not limited to a single syntactic position (Matisoff, 1969). A word’s freedom to occur in the N-position as well as in the V-position is one instance of versatility. In the extreme case of Late Archaic Chinese (5th–3rd century BC), any lexical item can take the verbal position, even a proper name:

- (2) Late Archaic Chinese (Zuo, Ding 10)
 Gōng Ruò yuē: ěr Wú wáng wǒ hū?
Gong Ruo say you Wu king I Q
 ‘Gong Ruo said: “Do you want to deal with me as King Wu was dealt with?”’
 (King Wu was murdered. → “Do you want to kill me?”)

Versatility may also enhance grammaticalization in the sense that full lexemes can take positions associated with grammatical functions. As is typical of versatility, one lexical item can take different functions depending on the construction in which it is used. Thus, the verb *ʔaoy* ‘give’ in Khmer can occur as a coverb (3), as a causative verb (4), or as an adverbial subordinator (5). The same applies to Vietnamese *cho* ‘give’ and to Thai *hây* ‘give’ (cf. Bisang, 1996: 577–578).

- (3) kəət baək tvi:ə(r) ʔaoy khjom
he open door give.COVERB I
 ‘he opens the door for me’
- (4) mda:y-mi:ŋ sovaŋ(ŋ) ʔaoy sva:myy cù:n
aunt Sovan give.CAUS husband give a lift
 phniəv tɿu phtəəh
guest Vd:go house/home
 ‘Aunt Sovan had her husband bring the guests back home’ (Bisang, 1992: 440)
- (5) khjom khom thvə:ka:(r) ʔaoy
I try hard work so that.COMP
 ʔo:pük khjom səpba:y-cvt(t)
father I be-pleased
 ‘I am working hard so that my father will be pleased.’

The versatility of lexical items may turn out to be another consequence of the high relevance of pragmatics in the sense that the positioning of lexical items into syntax is governed to a lesser or to a greater degree by pragmatics.

Some Individual Structural Properties of the Languages of Mainland Southeast Asia

This section will mostly refer to one or more of the following languages: Chinese (Mandarin Chinese) (Sinitic), White Hmong (Hmong Daw) (Hmong-Mien), Vietnamese (Mon-Khmer), Thai (Tai) and Khmer (Mon-Khmer). Apart from word order; classifiers; directional verbs (Vd), coverbs (COV), and TAM markers derived from verbs; and different functions of the verb ‘come to have,’ there are other characteristics of mainland Southeast Asia as a zone of convergence which will not be discussed here. I would like to refer to relational nouns (nouns such as Thai *nâa* ‘front’ in adpositional function with the meaning ‘in front of’), causatives marked by the verbs ‘make, do’ and ‘give, allow’ in Vietnamese, Thai, and Khmer (Bisang, 1992: 42–44; see also example (4) above), passivelike constructions with a tendency to adversative meaning, complementizers and adverbial subordinators derived from verbs such as ‘say,’ ‘give’ (cf. example (5) above), ‘finish’ and others, and,

Table 5 Word order in mainland southeast Asia

| | <i>Verb/Object</i> | <i>Adposition</i> | <i>Demonstrative</i> | <i>Classifier</i> | <i>Possessor/Genitive</i> | <i>Relative Clause</i> |
|------------|--------------------|-------------------|----------------------|-------------------|---------------------------|------------------------|
| Chinese | VO | Prep/Postp | DemN | CIN | GenN | RelN |
| Hmong | VO | Prep | NDem | CIN | NGen | NRel |
| Vietnamese | VO | Prep | NDem | CIN | NGen | NRel |
| Thai | VO | Prep | NDem | NCI | NGen | NRel |
| Khmer | VO | Prep | NDem | NCI | NGen | NRel |

finally, comparative constructions based on verbs with the meaning ‘surpass’ (e.g., in Cantonese, Thai, Vietnamese).

Word Order

The large majority of the languages belonging to the mainland Southeast Asian convergence zone are VO (verb middle, including Chinese [Mandarin Chinese]). The noun phrase is subject to variance. While Chinese (Mandarin Chinese) is consistently head final, Thai and Khmer are consistently head initial. Hmong and Vietnamese are head initial with the exception of the classifier phrase, which follows the Chinese (Mandarin Chinese) example. There are prepositions (coverbs) in all the languages; Chinese also has postpositions (relational nouns, i.e., nouns in adpositional function). Since numerals covary with classifiers, there is no extra column for them in Table 5.

Classifiers

Classifiers are minimally used with numerals, where their presence is overwhelmingly compulsory in mainland Southeast Asian languages. Thus, a Chinese (Mandarin Chinese) noun like *xìn* ‘letter’ must take a classifier (*fēng*) if it is counted:

- (6) sān fēng xìn
 three CLASS letter
 ‘three letters’

There is an implicational correlation between the existence of a classifier and the lack of obligatory number distinction (transnumerality): “Numeral classifier languages generally do not have compulsory expression of nominal plurality, but at most facultative expression” (Greenberg, 1974: 25). Since nouns in mainland Southeast Asian languages only denote a concept without any commitment to number, one of the functions of the classifier is to make that concept accessible by individuating it, i.e., by highlighting one of its conceptual boundaries which qualify it as a unit (Bisang, 1999). The semantic criteria for highlighting a concept also classify that concept. Typical criteria for classification are material (animate,

abstract, inanimate), shape (one-/two-/three-dimensional), consistency (flexible, hard or rigid, discrete), size (big, small), location (classifiers for plots of land, countries, gardens, fields, etc.) and spatial arrangement (Allan, 1977). Other criteria are based on physical, functional, and social interaction with the concept to be classified (Denny, 1976).

Classification can not only be used to individuate a concept by highlighting some of its properties; it can also be used for identifying one or more relevant objects denoted by a concept. While identification can take place without referring to individuation – one can identify an ‘apple’ without referring to its conceptual boundaries – it seems difficult to individuate it without simultaneously identifying it. Departing from classification, one can thus establish the following hierarchy:

- (7) classification > identification > individuation

Identification can be used either to mark the definiteness or specificity of a concept (referentialization) or to make it accessible for construction with for example a possessor or a relative clause (relationalization).

Taking together the functional range of classifiers in mainland Southeast Asia, there are no fewer than four functions: classification, individuation, referentialization, and relationalization. These functions are not equally distributed across Southeast Asia. The minimal functions operating in all the languages are classification and identification. Table 6 provides a survey (Bisang, 1999).

The following examples from Hmong illustrate the functions of classification/individuation (8a) and of relationalization (possession) (8b).

- (8a) peb rab riam
 three CLASS knife
 ‘three knives’
 (8b) nws rab riam
 he CLASS knife
 ‘his knife’

The referential function of classifiers is more difficult to show because this needs a lot of text. Once a concept is introduced, it can be marked as definite, sometimes by the classifier alone, sometimes by classifier plus demonstrative, sometimes only by the

demonstrative (for more, see Bisang, 1999: 152–153). It is, however, necessary to point out that reference marking is not compulsory. Thus, an unmarked noun can get any possible referential interpretation depending on context.

Directional Verbs, Coverbs and TAM Markers, and Syntactic Patterns

Markers derived from lexical items used for expressing directionality, adpositional functions, and tense-aspect-modality (TAM) are widespread in mainland Southeast Asia.

The direction taken by a state of affairs can be overtly expressed with directional verbs (Vd). Verbs belonging to this category have the meanings of ‘come,’ ‘go,’ ‘move upwards,’ ‘move downwards,’ ‘move into,’ and ‘move out of.’ In Khmer, there is a maximum of three slots (9), while Thai has only two slots and Vietnamese only one.

- (9) kəət yək ʔyvan coh
 he take luggage move. down.DIR
 cəŋ mək.
 move. out.DIR come.DIR
 ‘He takes [his] luggage down and out [of his
 room upstairs towards the speaker]’

Verbal lexemes in adpositional function are called coverbs (COV) (see Clark, 1978). An instance of the verb ‘give’ in that function is discussed in the subsection ‘Versatility.’ Other frequently used verbs

are ‘be at’ (locatives or directionals), ‘arrive’ (directionals), ‘move along something’ (path), ‘use’ (instrumental), ‘be equal to’ (work as, do something in the function of), and ‘replace’ (instead of). The following example is from Vietnamese:

- (10) tôi làm việc ở Sài Gòn.
 I do work be.at.COV Saigon
 ‘I work in Saigon’

Verbs in the function of TAM markers can occur in the preverbal position or clause finally (in Chinese [Mandarin Chinese], there are also the three TAM markers *-le*, *-zhe* and *-guo*, which are suffixed to the verb). The verb ‘finish’ in clause-final position, which is very widespread, is briefly looked at in this subsection (see also the next subsection on ‘come to have’). In Chinese (Mandarin Chinese), the clause-final marker *le* (derived from *liǎo* ‘finish’) marks a wide range of functions from perfect to the pragmatic function of reference to a preconstructed domain (Li *et al.*, 1982; Bisang and Sonaiya, 1997). Thai *léew*, which is borrowed from Chinese *liǎo* ‘finish,’ is an aspectual marker highlighting event-initial or event-final temporal boundaries. The functions of Hmong *lawm* (again related to Chinese *liǎo*) or *tas* (*lawm*) ‘finish,’ Vietnamese *rôi* ‘finish,’ and Khmer *haəy* (nowadays only used as a TAM marker, but cf. its transitive form *bəŋhaəy* ‘finish’) cover the same functional range as *le* and *léew*. Unfortunately, there is no detailed comparative analysis available.

If directional verbs, coverbs, and TAM markers are part of the same state of affairs, they follow a fixed pattern of word order (cf. serial unit in Bisang, 2001) described in Table 7 and illustrated by example (11) from Khmer.

- (11) kəət ba:n yək ʔyvan coh
 the be.able.TAM take luggage move.down.DIR
 cəŋ mək. ʔaoy khpom.
 move.out.DIR come. DIR give.COV I
 ‘he was able to bring [his] luggage down and out
 [of his room upstairs towards the speaker]
 to/for me.’

As can be seen from Table 7, the structure of the serial unit follows a certain areal clustering. Vietnamese, Thai, and Khmer, in the south, follow exactly the same pattern. Chinese (Mandarin Chinese), in the

Table 6 Functions of classifiers in individual languages

| | |
|--|--|
| I. <i>Classification & individualization</i> | |
| Modern Standard Chinese (Mandarin Chinese) (classifiers with numerals and demonstratives) | |
| Vietnamese (individualization, but not necessarily in the context of counting) | |
| II. <i>Classification & individualization & referentialization</i> | |
| Thai (secondary function in combination with stative verbs in N-CLASS-ADJ) | |
| III. <i>Classification & individualization & relationalization</i> | |
| Cantonese (yue Chinese) (classifiers can be used in possessive and relative constructions) | |
| IV. <i>Classification & individualization & referentialization & relationalization</i> | |
| Hmong (with referentialization being a secondary function) | |

Table 7 Positions within the serial unit

| Chinese | TAM | COV | V-TAM | COV | Vd | TAM |
|------------|-----|-----|-------|-----|-----|-----|
| Hmong | TAM | COV | V TAM | Vd | COV | TAM |
| Vietnamese | TAM | | V | Vd | COV | TAM |
| Thai | TAM | | V | Vd | COV | TAM |
| Khmer | TAM | | V | Vd | COV | TAM |

north, differs with regard to the following three properties: preverbal coverbs, TAM markers immediately after the verb and cov-Vd word order. Hmong lies in between. It shares the former two word order properties with Chinese (Mandarin Chinese) and the last property with the southern languages.

The above word order is not arbitrary even if one looks at languages spoken outside mainland Southeast Asia with comparable structures (Bisang, 2001: 202–214). It can be accounted for in terms of semantic generality as introduced by Bybee (1985). Increasing semantic generality of a marker is related to compatibility with more lexical stems and to greater morpho-syntactic fusion with the stem. Thus, maximally general grammatical categories are prototypically expressed inflectionally. Although there is no iconic correlation between the degree of semantic abstraction and morphological attrition in mainland Southeast Asia (see ‘Syllabic Morphology’ above), there is a form–meaning iconicity if one looks at the relative distance of the markers to the main verb. The further away a marker is from the main verb the more general it is. Directional verbs and coverbs still have enough semantic weight to be incompatible with many verbs. This is not the case with the semantically more general TAM markers. Therefore, TAM markers are situated at a greater distance from the verb than coverbs and directional verbs. Coverbs and directional verbs seem to share about the same degree of generality. Consequently, we find cov-Vd as well as Vd-cov.

The Verb ‘Come to Have’

There is an excellent study on the grammaticalization of the verb ‘get, come to have’ from an areal perspective in mainland Southeast Asia in Enfield (2003). Verbs such as Chinese (Mandarin Chinese) *dé*, Hmong *tau*, Vietnamese *đuợc*, Thai *dây*, or Khmer *ba:n* with that meaning induce a large number of different inferences depending on the context. Enfield (2003) translates these verbs with ‘come to have.’ This translation is more adequate than the one with ‘get’ because it does not imply an agentive subject.

Although ‘come to have’ verbs occur preverbally as well as postverbally, a look in this subsection at their preverbal functions will be enough to illustrate the rich inferential potential. Their basic meaning in this position is that the state of affairs expressed by the main verb is true or applies “because of something else that happened before this” (Enfield, 2003: 292). A set of inferences depends on whether the state of affairs denoted by the verb is understood as [+wanted] or [–wanted]. If it is wanted, we get either an abilitative (be able) or a permissive (be allowed)

inference (12); if it is not wanted, we get a strong deontic (must, have to) interpretation (13).

- (12) Hmong (Mottin, 1980; Bisang, 1992: 241):
 koj mus deev hluar nkauj,
 you go court girl
 koj puas tau nrog tham?
 you Q PFV with.COV talk
 ‘you courted the girl, did you [manage to] talk to her?’
- (13) Chinese (Mandarin Chinese) (in its deontic function, *dé* ‘get’ becomes *děi*):
 tā děi xuéxi zhōngwén.
 s/he must learn Chinese
 ‘S/he must learn Chinese’

Many grammarians of individual languages describe ‘come to have’ verbs as past markers. In spite of this, past is only another possible inference, but not a clear-cut grammatical category. The following Khmer example from Enfield (2003: 314) can also trigger other temporal inferences in other contexts:

- (14) khnom ba:n riəp-ka:(r)
 I get.TAM marry
 ta:m prəpəyŋi: khmae(r)
 according.to.COV custom Khmer
 ‘I married according to Khmer custom’
 (could mean in other contexts ‘I would/will get to/have to marry . . .’)

A fourth inference, treated only marginally by Enfield (2003), is emphasis of the truth. This inference is related to the fact that for the agent to be able to ‘come to have’ a given state of affairs, that state of affairs needs to be true.

- (15) Hmong (Mottin, 1980: 94):
 saib yog leej twg tau
 look be man/CL which PFV
 ua txhaum zoo li cas lawm?
 make mistake like.this PERF
 ‘[we want to] see who [really] made such a mistake’

In spite of their rich inferential potential, most languages show preferred inferences or even completely exclude certain inferences. Thus, the meaning of preverbal ‘come to have’ in Chinese (Mandarin Chinese) is conventionalized into deontic modality. Vietnamese prefers the abilitative or permissive interpretation but is compatible with a must-interpretation. Thai and Khmer are more open, with a functional core of abilitative/permissive and past. Hmong shows a certain preference for past (Enfield, 2003: 319) but certainly does not exclude abilitative/permissive inferences.

Conclusion – Factors Leading to a Zone of Convergence

Mainland Southeast Asia is characterized by a special type of pragmatics-oriented grammaticalization and by a number of shared products of grammaticalization and syntactic patterns. The structural convergence observed in this zone is the result of a very complex interaction of cognitive (semantic and pragmatic) factors with social mechanisms of diffusion extended over a large number of different individual situations of contact. Language-internal cognition-based processes of change are combined with and sometimes enhanced or interrupted by contact-induced changes. When it comes to social factors, any of the social models accounting for the cross-linguistic diffusion of structural properties presented in the literature can contribute their part. Thus, the diffusional pattern of the properties relevant for mainland Southeast Asia as a zone of convergence is most likely a joint product of social networks, leaders of linguistic change, and invisible-hand processes.

Bibliography

- Allan K (1977). 'Classifiers.' *Language* 53, 285–311.
- Bisang W (1992). *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer. Vergleichende Grammatik im Rahmen der Verbserialisierung, der Grammatikalisierung und der Attraktorpositionen*. Tübingen: Gunter Narr.
- Bisang W (1996). 'Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages.' *Studies in Language* 20(3), 519–597.
- Bisang W (1999). 'Classifiers in East and Southeast Asian languages: counting and beyond.' In Gvozdanović J (ed.) *Numeral types and changes worldwide*. Berlin: Mouton de Gruyter. 113–185.
- Bisang W (2001). 'Areality, grammaticalization and language typology: on the explanatory power of functional criteria and the status of universal grammar.' In Bisang W (ed.) *Aspects of typology and universals*. Berlin: Akademie-Verlag. 175–223.
- Bisang W & Sonaiya R (1997). 'Perfect and beyond, from pragmatic relevance to perfect: the Chinese sentence particle *le* and Yoruba *ti*.' *Sprachtypologie und Universalienforschung* 50(2), 143–158.
- Bybee J (1985). *Morphology: a study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Capell A (1979). 'Further typological studies in Southeast Asian languages.' In Nguyen D L (ed.) *Southeast Asian Linguistic Studies*, vol. 3. Canberra: Australian National University. 1–42.
- Chappell H (ed.) (2001). *Sinitic grammar: synchronic and diachronic perspectives*. Oxford: Oxford University Press.
- Clark M (1978). *Converbs and case in Vietnamese*. Canberra: Australian National University.
- Clark M (1989). 'Hmong and areal Southeast Asia.' In Bradley D (ed.) *Southeast Asian syntax*. Canberra: Australian National University. 175–230.
- Denny P J (1976). 'What are noun classifiers good for?' *Papers from the 12th International Meeting, Chicago Linguistic Society*, 122–132.
- Diffloth G & Zide N (1992). 'Austroasiatic languages.' In Bright W (ed.) *International encyclopedia of linguistics (4 vols)*, vol. 1. New York: Oxford University Press. 137–142.
- Diller AV N (1988). 'Thai syntax and "national grammar."' *Language Sciences* 10(2), 273–312.
- Edmondson J A & Solnit D B (eds.) (1997). *Comparative Kadai: The Tai branch*. Dallas: Summer Institute of Linguistics.
- Enfield N J (2003). *Linguistic epidemiology: semantics and grammar of language contact in mainland Southeast Asia*. London & New York: RoutledgeCurzon.
- Greenberg J (1974). 'Numeral classifiers and substantival number: problems in the genesis of a linguistic type.' In *Proceedings of the 11th International Congress of Linguistics, Bologna–Florence, Aug–Sept 1972*. 17–37.
- Haiman J (1998). 'Possible origins of infixation in Khmer.' *Studies in Language* 22(3), 597–617.
- Huang C-T J (1984). 'On the distribution and reference of empty pronouns.' *Linguistic Inquiry* 15, 531–574.
- Huang Y (1994). *The syntax and pragmatics of anaphora: a study with special reference to anaphora*. Cambridge: Cambridge University Press.
- Huffman F E (1973). 'Thai and Cambodian: a case of syntactic borrowing?' *Journal of the American Oriental Society* 93(4), 488–509.
- Huffman F E (1986). *Bibliography and index of mainland Southeast Asian languages and linguistics*. New Haven & London: Yale University Press.
- Jenner P N & Pou S (1980–1981). *A lexicon of Khmer morphology*. Honolulu: University of Hawai'i Press.
- Lehmann C (1995). *Thoughts on grammaticalization*. Munich: Lincom Europa.
- Li F (1977). *A handbook of comparative Tai*. Honolulu: University of Hawai'i Press.
- Li C N, Thompson S A & Thompson R M (1982). 'The discourse motivation for the perfect aspect: the Mandarin particle *le*.' In Hopper P J (ed.) *Tense-Aspect: between semantics and pragmatics*. Amsterdam & Philadelphia: John Benjamins. 19–44.
- Matisoff J A (1969). 'Verb concatenation in Lahu: the syntax and semantics of "simple" juxtaposition.' *Acta Linguistica Hafniensia* 12(1), 69–120.
- Matisoff J A (1991). 'Areal and universal dimensions of grammaticization in Lahu.' In Traugott E C & Heine B (eds.) *Approaches to grammaticalization*. Amsterdam: John Benjamins. 383–453.
- Mottin J (1980). *Contes et légendes Hmong Blanc*. Bangkok: Don Bosco.
- Wyatt D K (1982). *Thailand: a short history*. New Haven & London: Yale University Press.

Southern Bantu Languages

L Marten, School of Oriental and African Studies,
London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Southern Bantu languages include Bantu languages spoken in South Africa, Swaziland, Lesotho, Botswana, Zimbabwe, and southern Mozambique. The term 'southern Bantu languages' is usually taken to refer to the geographical-referential classification, and hence as not implying genetic relations, of the following languages and language groups: the Nguni group (including Zulu, Xhosa, Swati, Ndebele), the Sotho-Tswana group (including Northern Sotho, Sesotho [Sotho, Southern], Tswana), the Tswa-Ronga group, the Imhambane group, and also Shona and Venda. The Bantu languages of Angola and Namibia, such as Herero or Wambo, are usually not included under southern Bantu, but are referred to as south-western Bantu. In terms of Guthrie's (1967–1971) classification, southern Bantu languages are grouped as zone S. The designation 'southern Bantu languages' was re-enforced by Doke's monograph with the same title published in 1954.

Historically, the southern Bantu languages provide the endpoint of the so-called Bantu expansion, a period of migration and contact of more than 2000 years, during which Bantu languages slowly came to be spoken throughout the larger part of sub-Saharan Africa. The origin of the Bantu expansion lies in the Nigeria-Cameroon borderland, and the direction of the expansion was hence southwards and ended with Bantu languages reaching the southern African coastline about 1,500 years ago, coming from eastern and central Africa. Speakers of southern Bantu languages have probably shared an extended, and extensive, period of contact with speakers of Khoisan languages present in southern Africa when Bantu languages arrived. The differentiation of distinct Bantu languages, and the establishment of standardized forms occurred more recently. From the 19th century onwards, written literature was produced in the larger southern Bantu languages, and several are among the dominant languages in the countries where they are spoken. For example, all nine Bantu languages of the national languages of the Republic of South Africa (in other words, all national languages except for English and Afrikaans) are southern Bantu languages.

Southern Bantu languages have played an important part in the history of Bantu studies. While the earliest descriptions of Bantu languages are from

west-central and eastern Africa, scholarship in southern African Bantu languages provided the impetus for a number of early comparative Bantu studies (Lichtenstein, 1808; Bleek, 1862, 1869; Torrend, 1891). Bleek (1862) is credited with coining the term 'Bantu' based on the plural form for 'people' in Xhosa and in many other Bantu languages. In the 20th century, a major figure in the study of southern Bantu languages was Clement Doke, who produced numerous grammatical descriptions of southern Bantu languages and also proposed an analytical system for the description of Bantu languages, which became particularly influential in South Africa (Doke, 1935). Due to the official status of nine southern Bantu languages in South Africa and an increase in institutions of tertiary education, there is currently a wealth of new linguistic scholarship in southern Bantu languages, often with particular emphasis on lexicography, computational linguistics, and applied linguistic topics such as language policy and language teaching. Most of the southern Bantu languages with large numbers of speakers (Zulu, Xhosa, Northern Sotho, Sesotho, Tswana) have, often recent, comprehensive reference grammars and dictionaries, as well as a range of teaching materials for schools and independent learners.

Classification

The southern Bantu languages are, following Guthrie (1967–1971), classified into six groups within zone S (cf. Gowlett, 2003).

The Shona group (S10) comprises six clusters: Korekore, Zezuru, Manyika, Karanga, Nda, and Kalanga. A standardized form of Shona based on the Korekore, Zezuru, and Karanga varieties is used as an official language in Zimbabwe. In addition to Zimbabwe, languages of the Shona group are also spoken in parts of Mozambique and Botswana (Kalanga). There are close to 9 million speakers of Shona.

The Venda group (S20) only includes Venda, spoken by around 800 000 speakers in South Africa's Northern Province and adjacent southern Zimbabwe. Venda is an official language of South Africa.

The Sotho-Tswana group (S30) includes Tswana, Northern Sotho, and Southern Sotho, all of which are cover terms for a number of related varieties. Sometimes also Lozi, spoken in western Zambia and Namibia's Caprivi strip, is classified as a Sotho-Tswana language. Standard forms of these languages are based on a majority variety, and smaller varieties are often threatened with marginalization. Tswana is an official language in Botswana and South Africa,

with about 4 million speakers. Northern Sotho, also Sesotho sa Leboa, is spoken in the northeast of South Africa by about 3.6 million speakers and is an official language of South Africa. Southern Sotho, or Sesotho, with more than 4 million speakers, is spoken in Lesotho and South Africa and it is an official language in both countries.

The Nguni group (S40) is divided into Zunda varieties and Tekela varieties. Among the Zunda varieties are Xhosa, Zulu, and Zimbabwean Ndebele. Xhosa includes a number of different varieties. Zulu, with around 10.7 million speakers, and Xhosa, with around 7.2 million speakers, are official languages of South Africa. Zimbabwean Ndebele has official status in Zimbabwe. The Tekela varieties include Swati, South African Ndebele, and the smaller languages Phuthi and Lala (Lala-Bisa). Swati has around 1.6 million speakers and is an official language both in Swaziland and South Africa. The southern variety of South African Ndebele is an official language in South Africa, spoken by around 0.6 million speakers.

The Tshwa-Ronga group (S50) includes Tshwa, Tsonga, and Ronga, all of which are spoken in Mozambique. Tshwa, with 0.7 million speakers, is also spoken in Zimbabwe, and Tsonga, with more than 3 million speakers, is also spoken in South Africa, where it is an official language.

The Inhambane, or Copi, group (S60) includes two languages spoken in the Inhamabane area of Mozambique: Copi (Chopi), with around 0.5 million speakers, and GiTonga, with around 0.3 million speakers.

Structural Features

Phonologically, southern Bantu languages are characterized by symmetric five (e.g., Nguni) or nine (e.g., Sotho) vowel systems, two tonal distinctions (high vs. low), and complex consonant systems, often including a three-way distinction between voiceless-aspirated, voiceless-unaspirated, and voiced stops and affricates, as well as several series of pre-nasalized consonants. A number of southern Bantu languages have borrowed click consonants from Khoisan languages during an extended period of contact, e.g., Xhosa, which has dental [ǀ], alveolar [ǃ] and lateral [ǁ] clicks (written as *c*, *q*, and *x*). A comparatively untypical Bantu feature of southern Bantu languages are depressor consonants, an often phonologically heterogeneous group of consonants that cause a following high tone to lower, as in the Zulu example below, where the depressor consonant /z/ in the plural prefix causes the following high tone to shift to the following syllable, resulting in a different tone pattern:

- (1) *isihlālò* 'chair' *izihlālò* 'chairs'

Morphologically, southern Bantu languages, like the majority of Bantu languages, are characterized by their noun classes, and the agreement, or concord, system built on it, as well as by complex verbal morphology. Nouns are grouped into 15 to 20 noun classes that are morphologically marked by a noun class prefix, usually of CV shape and sometimes accompanied by a pre-prefix vowel (see Table 1).

The noun class of the head noun triggers class agreement of dependent nominals, as well as subject and object concord (agreement) morphology in the inflected verb. The term agreement, although well established, can be misleading, as subject and object markers can function as subject and object, and no overt lexical NP is needed for a well-formed sentence, as the following Zulu example (from Poulos and Bosch, 1997) shows:

- (2) *ngi-zo-ba-sebenz-el-a*
SM1sg-FUT-OM2-work-APPL-FIN
'I will work for them'

In addition to subject and object markers, inflected verbs can show morphological marking of negation, tense, aspect and mood, typically prefixed to the verbal base. The verbal base consists of a root that may be suffixed by several derivational suffixes (so-called extensions), such as applicative, causative, stative, reciprocal, or passive. The following examples from Tswana show how causative and applicative extensions can be used to increase the number of nominal

Table 1 Noun class prefixes in southern Bantu languages

| Class | Shona | Venda | Sesotho | Zulu | Tsonga | Copi |
|-------|-------|-------|---------|--------|--------|--------|
| 1 | mu- | mu- | mo- | um(u)- | mu- | in- |
| 2 | va- | vha- | ba- | aba- | va- | va- |
| 3 | mu- | mu- | mo- | um(u)- | mu- | in- |
| 4 | mi- | mi- | me- | imi- | mi- | mi- |
| 5 | Ø- | ǀi- | le- | i(li)- | ri- | di- |
| 6 | ma- | ma- | ma- | ama- | ma- | ma- |
| 7 | chi- | tshi- | se- | isi- | xi- | tshi- |
| 8 | zvi- | zwi- | di- | izi- | svi- | si- |
| 9 | N- | N- | N- | iN- | N- | (N-) |
| 10 | N- | dziN- | di(N)- | iziN- | ti(N)- | ti(N)- |
| 11 | ru- | lu- | - | u(lu)- | ri- | li- |
| 12 | ka- | - | - | - | - | - |
| 13 | tu- | - | - | - | - | - |
| 14 | (h)u- | vhu- | bo- | ubu- | vu- | wu- |
| 15 | ku- | u- | go- | uku- | ku- | ku- |
| 16 | pa- | fhu- | fa- | (pha-) | ha- | ha- |
| 17 | ku- | ku- | go- | (ku-) | ku- | - |
| 18 | mu- | mu- | mo- | - | mu- | - |
| 19 | svi- | - | - | - | - | - |
| 20 | - | ku- | - | - | - | - |
| 21 | zi- | ǀi- | - | - | ji- | - |

complements of the verb (adapted from Creissels, 2004):

(3a) ng-wàná ó nó-lé má-ši
NP1-*child* SM1 *drink*-PERF NP6-*milk*
'the child drank milk'

(3b) kè nó-s-ítsé ng-wàná má-ši
SM1sg. *drink*-CAUS- NP1- NP6-
 PERF *child* *milk*
'I made the child drink milk'

(3c) kè nó-s-éd-ítsé Dimpho
SM1sg. *drink*-CAUS-APPL-PERF *Dimpho*
ng-wàná má-ši
NP1-*child* NP6-*milk*
'I made the child drink milk in Dimpho's
place'

A number of morphological features found in southern Bantu languages are not, or only rarely, found in other Bantu languages. In the domain of nominal morphology, these include the use of derivational suffixes for diminutives and feminines (e.g., in Zulu *-ana* and *-kazi*: *indoda*, 'man' and *indodana*, 'son'; *imbuzi* 'goat' and *imbuzikazi*, 'she-goat'), and the replacement of the locative classes by a locative prefix *e-*, a locative suffix *-(i)ni*, or a combination of both. Within verbal inflection, several southern Bantu languages show a distinction between so-called conjoint and disjoint verb forms (sometimes also called definite/indefinite or long/short forms) (Creissels, 1996); for example, in Zulu, verbs in the perfect tense may end in *-e* (short) or *-ile* (long) (Doke, 1963: 335):

(4a) si-bon-e abantu
SM1pl-*see*-PERF *people*
'we saw people'

(4b) si-ba-bon-ile abantu
SM1pl-OM2-*see*-PERF *people*
'we saw (them) the people'

The difference in use of the two forms depends on different factors, among them whether the verb is final in the verb phrase, as in (4b), where the object marker functions as the object of the verb, and the overt NP *abantu* following the verb is hence not part of the VP.

In terms of syntax, southern Bantu languages, like most Bantu languages, have unmarked SVO order (see the examples in (3), above), but, especially in interaction with subject and object markers, the word order is syntactically comparatively free, and rather more constrained by information structure considerations. As an illustration of this, the following Xhosa examples show unmarked subject-verb order (5) and inverted verb-subject order (6) used to focus the subject *ábántwánà*, 'children', either existentially like in this example, or contrastively, as in

(7). Note that the subject marker in (6) and (7) is of the locative class 17 and thus does not show agreement with the subject (Du Plessis and Visser, 1992: 130–131):

(5) ábá-ntwánà bá-yà-ngénà
NP2-*children* SM2-*ya-enter*
'the children enter'

(6) kù-ngénà ábá-ntwánà
NP2-*children* SM2-*ya-enter*
'there enter children'

(7) kù-sèbénzà áamá-dódà, háyi ábá-fázi
SM17-*work* NP6-*men* *not* NP2-*women*
'there are men working, not women'

Although the major southern Bantu languages, those with large numbers of speakers and official status, are comparatively well-described, for a number of smaller southern Bantu languages, some of which are endangered, very little information exists, and descriptive studies are urgently needed. In a wider perspective, the contribution southern Bantu languages can make to general and theoretical linguistic studies has only begun to be fully addressed, and remains to be developed in all areas of linguistic research in the future.

Bibliography

- Bleek W H I (1862). *A comparative grammar of South African languages 1: Phonology*. Cape Town: J. C. Juta/London: Trübner & Co.
- Bleek W H I (1869). *A comparative grammar of South African languages 1: The concord 1: The noun*. Cape Town: J. C. Juta/London: Trübner & Co.
- Brauner S (1995). *A grammatical sketch of Shona*. Köln: Köppe.
- Cole D T (1955). *An introduction to Tswana grammar*. Cape Town: Longman.
- Creissels D (1996). 'Conjunctive and disjunctive verb forms in Setswana.' *South African Journal of African Languages* 16, 109–114.
- Creissels D (2004). Non-canonical applicatives and focalization in Tswana, to appear in *Syntax of the World's Languages*.
- Doke C (1935). *Bantu linguistic terminology*. London: Longman.
- Doke C (1954). *The southern Bantu languages*. London: Oxford University Press for the International African Institute.
- Doke C (1963). *Textbook of Zulu grammar* (6th edn.). Cape Town: Maskew, Miller, Longman.
- Du Plessis J A & Visser M (1992). *Xhosa syntax*. Pretoria: Via Afrika.
- Fortune G (1955). *An analytical grammar of Shona*. Cape Town: Longman.
- Gowlett D (2003). 'Zone S.' In Nurse D & Philippson G (eds.) *The Bantu languages*. London: Routledge. 609–638.

- Guthrie M (1967–1971). *Comparative Bantu* (4 vols). Farnborough: Gregg.
- Joffe D (2004). *African Languages*. Internet site. <http://africanlanguages.com>. Accessed 18 Sept 2004.
- Lichtenstein M H K (1808). 'Bemerkungen über die Sprachen der südafrikanischen wilden Völkerstämme.' *Allgemeines Archiv für Ethnographie und Linguistik* 1, 259–331.
- Olivier J (2004). South African Languages Web. Internet site. <http://salanguages.com>. Accessed 18 September 2004.
- Poulos G (1990). *A linguistic analysis of Venda*. Pretoria: Via Afrika.
- Poulos G & Bosch S (1997). *Zulu*. München: Lincom.
- Poulos G & Louwrens L J (1994). *A linguistic analysis of Northern Sotho*. Pretoria: Via Afrika.
- Poulos G & Msimang C (1998). *A linguistic analysis of Zulu*. Pretoria: Via Afrika.
- Torrend J (1891). *A comparative grammar of the South African Bantu languages*. London: Kegan Paul, Trench, Trübner & Co.
- Statistics South Africa (2004). *Census 2001: Primary tables South Africa*. Pretoria: Statistics South Africa.
- Ziervogel D & Mabuza E J (1976). *A grammar of the Swati language (siSwati)*. Pretoria: van Schaik.

Spanish

R Wright, University of Liverpool, Liverpool, UK

© 2006 Elsevier Ltd. All rights reserved.

Spanish is the standard language of over 300 million people in Spain, Equatorial Guinea, and 18 states in Latin America; it is also widely used in the United States, Israel, and in Western (former Spanish) Sahara. The standard is based on, and almost identical with, the Romance speech of Old Castile, which is why non-Castilians tend to call it *castellano* rather than *español* and sometimes resent its privileged status; for according to the Spanish constitution, all Spaniards have the obligation to learn it and the right to use it, which has made both its use and its name serious and even dangerous political issues in areas where many people are native speakers of another language (see **Catalan**; **Basque**), or of the *babes* of Asturias.

History

The Romans came to Spain during the Punic Wars of the late third century B.C. Their language has been spoken in the Peninsula ever since (see **Latin**; **Romance Languages**). Iberian Romance languages only began to acquire separate names and identities in the 13th century; until then it is simplest to envisage one single though heterogeneous Romance speech community throughout the Peninsula. The Romance (*mozárabe*) of bilingual Arabic–Romance areas was probably barely influenced by Arabic and similar to that of Christian areas. The traditional writing techniques survived as the official written standard till the early 13th century, but the techniques used then in the first texts exclusively prepared in the new written form ('Old Spanish') are based on unofficial experimentations that can be traced from the 11th century.

Later in that century it was decided in the Kingdom of Castile (which included Leon and Galicia) to base their written standard on the speech of Castile. This written standard was extended to Aragon and Catalonia after the union of Spain in 1479, even though Aragonese and Catalan already had written standards of their own, and was the only written form exported to the New World. In the 18th century the newly founded Spanish Academy standardized written Castilian almost definitively. The spread of spoken Castilian is less easy to chart; many people fluently speak Catalan, Aragonese, Leonese, or Galician, but read or write only Castilian.

Phonetics and Phonology

Standard Castilian has 18 consonantal phonemes: /b, p, d, t, g, k, f, θ, s, x, tʃ, m, n, ɲ, l, ʎ, r/. /k/ has almost entirely delateralized to merge with /j/, often realized as [ʒ] or [dʒ]; word-initially it derives from /pl-/ or /kl-/ (e.g., Latin *clavem*, Spanish *llave* 'key'; cf., Portuguese *chave* ([ʃ-]), Italian *chiave* ([kj-]), French *clef*). Voiced plosives occur only breathgroup-initially or after nasal consonants, being so outnumbered by the fricative allophones used elsewhere that some linguists prefer to annotate the phonemes as /β, ð, γ/. Preconsonantal nasals are homorganic. /s/ is realized [z] before a voiced consonant. In most of Andalucía and all of America there is no distinction made between /θ/ and /s/; usually they merge as syllable-initial [s] and syllable-final [-h] (or ø), but in parts of rural Andalucía as [θ]. The phonemic status of the two semi-vowels /w/ and /j/ is controversial; they might be allophones of /u/ and /i/, respectively. There are only five vowels: /a, e, i, o, u/. Rising diphthongs are much more common than falling. Schwa is not found, but synalepha at word boundaries is normal (*diez y once* /djeθionθe/ [djéθjónθe] 'ten and eleven').

The preferred syllable structure is CV (ca., 56% of syllables), which overrides word boundaries (e.g., *cual es*, 'which is' [kwa + les]). Stress is largely predictable, given morphological information; many monosyllables are clitic. Intonation rarely varies more than an octave.

Morphology

The only nominal inflection is plural marking [-s] (postconsonantly [-es]). All nouns in use have to be either masculine or feminine gender, and adjectives display number and gender concord. There is an extensive system of verbal inflections; verbs are marked for number and person concord with their subjects. Several paradigms are in opposition according to mood, aspect, relative time, and subjective attitude, in ways still not entirely understood. The citation form is the infinitive, which always ends in a stressed theme vowel + [-r]. The majority end in [-ár], including all neologisms other than those with the inchoative affix *-ecer*; the rest end in [-ér] or [-ír], conjugations most of whose other inflections are shared. Second person singulars tend to end in [-s], first person plurals always in [-mos], and third person plurals always in [-n]. Several verbs have systematically patterned variation in their stems, e.g., stressed [je] versus unstressed [e] (*tener* [tenér] 'to have'; *tiene* [tjéne] 'he has'), or stem-final [θ] before front vowels versus [θk] before others (*conocer* [konoθér] 'to know'; *conozco* [konóθko] 'I know'). Irregular verbs usually belong to the [-er] or [-ir] category, combining irregular stems with regular inflections. Many verb forms employ auxiliaries, whose repertoire is numerous for progressives, while only *haber* is available for the perfect (thus *he venido pensando* 'I've been thinking'; *venir* 'come'); perfects are rarely used at all in northern Spain. Adverbs are formed off feminine adjectives with *-mente*. Derivational morphology is widely used; ostensible diminutives (*-illo*, *-ito*, and others) can be added to any nominal form with almost any meaning (depending on context and intonation); class-changing suffixes are used uninhibitedly (e.g., *-al* turns nouns into adjectives); meaningful prefixes are common, and the fashion for Verb + Plural direct object compounds with agentive meaning is spreading (e.g., *el tocadiscos*, literally 'the play-records,' 'the recordplayer').

Syntax

Sentences need no overt subject, e.g., *comíamos* 'we were eating,' *llueve* 'it's raining.' Some linguists unhelpfully postulate an underlying subject here. Adjectives follow nouns if clarifying the reference of the

NP, and precede it if the reference is already clear; if in doubt, listeners take the order to be NA. There is no general fixed order of verb and noun phrases; in general, the known precedes the unknown. Thus, *Juan llegó* 'John arrived,' if John has already been discussed, and *llegó Juan* if arrivals but not John have been discussed. SV order is never obligatory; VS is obligatory in *wh*-questions, outside the Caribbean, and normal in subordinate clauses (*la casa en que vivía mi madre* 'the house my mother lived in'). OV order is obligatory when clitic pronouns accompany finite verbs (*la vi*, 'I saw her'). A preposed nominal direct object requires a clitic copy, and in speech an indirect object in any position often has the same effect. Direct objects with particular reference, if misidentifiable otherwise as subjects, take a preposed *a* (*a la reina la vio, vio a la reina*, both 'he saw the Queen'); since *a* marks both direct/and indirect objects, and several Spanish speakers make no formal differentiation between direct and indirect object pronouns either, this direct/indirect distinction may be lapsing. The only preposition that can normally link nouns within a noun phrase is a correspondingly meaningless *de*. Articles are preposed: the so-called 'definite' article (*el, la, los, las*) is also used in generalizations; partitive use is often marked by the lack of any article.

The use of subjunctive or indicative mood is usually grammatically determined (e.g., *pido que* 'I ask for' is always followed by subjunctive), but the so-called 'past subjunctive' can also be used in subordinate clauses for already-known material; the 'past' subjunctive (which has two usually interchangeable paradigms) is in fact atemporal. Grammatically reflexive *se* is often used with passive or 'impersonal middle' meaning (*se abrió la puerta* 'the door (was) opened'); occasionally, in VS sentences of this type but not SV, a plural subject is preceded by a verb with singular concord (sometimes *se vende manzanas*; usually *se venden manzanas*; never **manzanas se vende*, 'apples for sale'), but this is nowhere the normal usage; linguists have tried and signally failed to analyze this *se* as a subject.

Vocabulary

The most startling fact about Spanish for an English speaker is the presence of two words for 'to be': *estar* (<Latin *stare* 'stand') and *ser* (suppletively < *sedere* 'sit,' and *esse* 'be'), only approximately distinguishable as being used for individual circumstances and general statements, respectively (and roughly vice versa when used as passive auxiliaries). Linguistic atlases, currently fashionable, show that lexical usage is noticeably not geographically standardized;

fish have different names, and the same word may be applied to different fish, in different parts, for example. Latin America has naturally adopted many local words of Indian provenance. Although the inherited vocabulary has been enriched by borrowings from Basque, Arabic, Catalan, French, Italian, Renaissance Latin, Nahuatl, Quechua, English, etc., most neologisms are more commonly formed via derivational morphology or semantic shift.

The Future

Spanish has wide geographical variation but remains a single speech community with a general standard for all to style-shift toward in formal situations, for the Latin-American standard is very similar to the European and will remain so, given mass communications. Local variations grow beneath the standards, however. For example, bilingual Aymara speakers in Bolivia have adopted the Aymara evidential system into their Spanish morphology, and Guaraní speakers in Paraguay have adopted Guaraní nominal tense-markers (e.g., *mi noviakué*, literally 'my girl-friend-past,' 'my former girlfriend'). Areas that aspirate or lose final /s/ have thereby lost a second person singular inflection and acquired homonymy with the third person forms and use subject *tú* more in compensation; the formal second person (third person morphology, with subject *usted(es)*) is anyway decaying in some places but strong in others, and the system varies greatly in America. The study of linguistics in Spanish universities is now flourishing, lively, and fashionable, putting Spain (temporarily, perhaps) in the vanguard of modern Romance linguistics. There

is still a great deal to discover and explain. For that reason a bibliography largely confined to English-language works is necessarily partial and parochial.

Bibliography

- Corominas J & Pascual J A (1980–1991). *Diccionario crítico etimológico castellano é hispánico* (6 vols). Madrid: Gredos.
- Del Valle J & Gabriel-Stheeman L (2002). *The battle over Spanish between 1800 and 2000: language ideologies and Hispanic intellectuals*. London: Routledge.
- Hualde J I et al. (2001). *Introducción a la lingüística española*. Cambridge: Cambridge University Press.
- Lipski J (1994). *Latin American Spanish*. London: Longman.
- Mackenzie I (2001). *A linguistic introduction to Spanish*. Munich: Lincom Europa.
- Mar-Molinero C (2000). *The politics of language in the Spanish-speaking world*. London: Routledge.
- Penny R (2000). *Variation and change in Spanish*. Cambridge: Cambridge University Press.
- Penny R (2002). *A history of the Spanish language* (2nd edn.). Cambridge: Cambridge University Press.
- Pountain C (2001). *A history of the Spanish language through texts*. London: Routledge.
- Pountain C (2003). *Exploring the Spanish language*. London: Arnold.
- Silva-Corvalán C (1989). *Sociolingüística: teoría y análisis*. Madrid: Alhambra.
- Stewart M (1999). *The Spanish language today*. London: Routledge.
- Tuten D N (2003). *Koineization in Medieval Spanish*. Berlin: Mouton de Gruyter.
- Whitley M S (2002). *Spanish/English contrasts* (2nd edn.). Washington, DC: Georgetown University Press.

Sumerian

G Cunningham, University of Oxford, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Sumerian – a long-dead language isolate documented throughout the Middle East, in particular in the south of what is now Iraq – rivals ancient Egyptian as the earliest written language. The first sources date to the late 4th millennium B.C.E. and the last to the 1st century C.E. When the language ceased to be spoken is uncertain: some estimates date this to the early 2nd millennium B.C.E. It was subsequently an elite language, used only in royal, ritual, and scholarly contexts. The language's users referred to it as *eme gir*,

possibly meaning 'native tongue', the term *Sumerian* being an anglicization of Akkadian *Šumeru*. The grammar outlined here is based on documents from the far south in the late 3rd millennium B.C.E., benefiting from unpublished work by Bram Jagersma, but it applies broadly to other places and periods.

Phonology

Fifteen consonants are used in transliterating Sumerian: <b, d, g, ḡ (/ŋ/ as in *sing*), ḥ (/x/ as in *loch*), k, l, m, n, p, r, s, š (/ʃ/ as in *ship*), t, z>. The language also had at least two weak consonants <y, ʔ (/ʔ/ glottal stop)>, which were subject to phonological change or loss in

certain environments, and eight vowel phonemes, short and long <a, e, i, u>. Neither vowel length nor weak consonants are indicated in transliteration. These alphabetic representations should be regarded as approximations. Vowel assimilation, both anticipatory and perseverative, is extensive, resulting in considerable allomorphic variation.

Word Classes

Nouns and verbs are the primary open word classes. In addition to numerals, the language has small closed classes of adjectives, adverbs, conjunctions, circumpositions, and interjections, as well as related sets of pronouns (personal, demonstrative, indefinite, interrogative, and reflexive) and determiners (possessive, demonstrative, and an indefinite). Most determiners cliticize to whatever class of word precedes them, as do plural marking and case marking. In the cuneiform script used to record Sumerian, lexical words are typically written logographically and all function morphemes (bases, clitics, and affixes) phonographically. Signs that constitute a word are linked by hyphens in transliteration, as are enclitics to their host. Given our uncertainty about the phonological form of many words, transliteration is simply a sign-by-sign representation of what was written.

Morphology

In terms of morpheme segmentation, Sumerian is more agglutinating than fusional. Inflectional affixation is restricted to verbs. Like verbs, nouns and adjectives can have reduplicated bases. Possibly, in nouns this expresses universality, in dynamic verbs iterativity, and in stative verbs intensity; its function in adjectives is unclear.

New nouns are mainly formed by compounding; new verbs are formed as by multiword expressions in which a noun and verb combine as a semantic unit, resulting in many three-place transitive predicates, such as:

ġiṣ tag
wood touch
'touch wood to something (i.e., sacrifice something)'

Nouns and Phrases

Nouns are marked for gender, although this distinction appears morphologically in only the pronominal morphemes and syntactically only in restrictions

that relate to plural and case marking. The gender distinction is between human (people and deities) and nonhuman (animals and inanimates), with some socially conditioned exceptions; in addition, non-human pronominal morphemes can refer to groups of people or deities. Only human nouns are marked for number.

At the level of the noun phrase, the language is left-headed, the sequence in outline being:

noun, modifier(s), determiner, plural marker, case marker

although the indefinite and most demonstrative determiners do not occur with modifiers.

Case markers typically indicate the syntactic role of the phrase in the clause. The core functions of the subject and direct object are marked by the ergative (*e* = transitive subject) and absolutive (\emptyset = intransitive subject and transitive direct object). Zero-marking is also used for personal pronouns as subject of both intransitive and transitive verbs. Noun phrases with a noun as head consequently follow ergative-absolutive alignment, whereas those with a personal pronoun as head follow nominative-accusative alignment. Personal pronouns occur infrequently (expressing emphasis or contrast); the language has person-number-gender (PNG) affixes in the finite verbal forms that index these core functions.

Table 1 shows the noncore adverbial case markers, arranged to reflect their relationship with a more nuanced set of morphemes incorporated in finite verbal forms. Like case markers, the verbal prefixes are postpositional in that they can be preceded by a noncore PNG prefix (see Table 2).

A further case marker, the genitive, is typically adnominal, marking noun phrase to noun phrase relations. It encompasses a much wider semantic field than possession. Genitive noun phrases may occur in the modifier(s) slot:

bad₃ iri kug-ga-ka-ni
bad iri kug = ak = ani = \emptyset
wall city holy = GEN = POSS.3.SING = ABS
'her wall in the Holy City'

(In this example, the first line is a transliteration of the script, in which the subscript numeral in *bad₃* is a modern convention that distinguishes between homophonous signs; the second line is a morphemic representation, in which = marks a clitic boundary.) This example indicates two characteristics of the script: (1) There is not always a one-to-one correspondence between morpheme and sign and (2) the reduplicated writings of consonants often appear to have no phonological implications.

Table 1 Sumerian noncore case morphemes

| Phrase-final enclitic | | Translation guide | Equivalent verbal prefix | |
|--|-----------------------|-----------------------|--------------------------|-------------------------------|
| Local | | | | |
| Dative (human only) | <i>r(a)</i> | 'to/for' | Dative | <i>a</i> |
| Directive (nonhuman only) ^a | <i>e</i> | | | |
| Dative (human only) | <i>r(a)</i> | 'in(to) contact with' | Directive ^a | <i>i</i> |
| Directive (nonhuman only) ^a | <i>e</i> | | | |
| Dative (human only) | <i>r(a)</i> | 'on(to)' | Directive ^a | <i>i</i> and <i>y (>e)</i> |
| Locative (nonhuman only) | ² <i>a</i> | | | |
| Locative (nonhuman only) | ² <i>a</i> | 'in(to)' | Locative (nonhuman only) | <i>n(i)</i> |
| Terminative | <i>š(e)</i> | 'to(ward)' | Terminative | <i>ši</i> |
| Ablative | <i>ta</i> | 'from' | Ablative | <i>ta</i> |
| Comitative | <i>d(a)</i> | '(together) with' | Comitative | <i>da</i> |
| Manner | | | | |
| Equative | <i>gin</i> | 'like' | None | |
| Adverbial | <i>eš</i> | 'in the manner of' | None | |

^aThe directive is sometimes referred to as the locative-terminative.

Table 2 Sumerian finite verb^a

| Extra-inflectional prefixes | Noncore prefixes | Core affixes and verbal base |
|-----------------------------|-------------------------|---------------------------------|
| <i>a(l)</i> or <i>i</i> | Noncore PNG | Core PNG |
| Clause | | |
| connective: <i>nga</i> | Dative | Verbal base |
| Cislocative: <i>m(u)</i> | Comitative | Aspect: <i>e(d)</i> or <i>Ø</i> |
| Middle passive: <i>ba</i> | Ablative or terminative | Core PNG |
| | Directive or locative | |

^aPNG = person-number-gender marker.

When genitive noun phrases express only possession, they have two characteristics: They are in complementary distribution with possessive determiners, and they can be shifted to the beginning of the clause, in which case a possessive determiner then is added to the original phrase (NHUM stands for nonhuman):

*e*₂-*a* *ni*₂
e = *ak* *ni*
 temple = GEN awesomeness
gal-*bi* *gal* = *bi* = *Ø* great = POSS.3.NHUM = ABS
 'the temple's great awesomeness'

Here the genitive is written *a*, the form used when it is not followed by a vowel.

Verbs and Clauses

Given that the dependents of a verb can be expressed in pronominal form by verbal affixes, a clause can consist only of a finite verb. However, in a clause that

includes noun phrases, the language is right-headed, the typical order being subject(–object)–verb.

A few verbs are irregular, having a different base depending on aspect and/or number; they can be divided into two major classes: reduplicating and suppletive. Plural bases are restricted to suppletive verbs; they are mostly used with a plural intransitive subject or direct object and thus follow ergative-absolutive alignment.

The Sumerian aspect and/or tense categories are difficult to reconstruct. Many Sumerologists have adopted instead two terms used by Akkadian grammarians, *hamtu* ('quick') and *marû* ('fat'). However, the principal distinction in finite verbal forms may be between completive and incomplete aspects (*hamtu* and *marû*, respectively). Nonfinite forms are more nuanced and have stronger temporal connotations, distinguishing between completive (typically with past reference), habitual (typically with present reference), and incomplete (typically with nonpast reference).

Stative verbs are excluded from incomplete aspect and only context indicates whether they have past or nonpast reference. In nonfinite verbal forms, intransitive stative verbs are in completive aspect and transitive ones are in habitual aspect. The copular verb is also excluded from incomplete aspect. It conjugates like an intransitive verb and is attested in both enclitic (prefixless) and independent forms.

Nonfinite verbal forms function as verbal adjectives and nouns, and in nonfinite relative and adverbial clauses (for example, of purpose and time). Their inflection comprises a prefix expressing negation, base reduplication, and an aspect suffix (²*a* = completive;

Table 3 Sumerian core (subject and direct object) affixes in finite verb

| | <i>Intransitive</i> | <i>Transitive</i> | | | |
|------------|-----------------------|------------------------------------|-----------------------------|-----------------------------|-----------------------|
| | <i>Both aspects</i> | <i>Completive aspect</i> | | <i>Incompletive aspect</i> | |
| | <i>Subject suffix</i> | <i>Subject prefix or circumfix</i> | <i>Direct object suffix</i> | <i>Direct object prefix</i> | <i>Subject suffix</i> |
| 1 HUM SING | <i>en</i> | 𒀭 | <i>en</i> | 𒀭 | <i>en</i> |
| 2 HUM SING | <i>en</i> | <i>y(>e)</i> | <i>en</i> | <i>y(>e)</i> | <i>en</i> |
| 3 HUM SING | Ø | <i>n</i> | Ø | <i>n</i> | <i>e</i> |
| 3 NHUM | Ø | <i>b</i> | Ø | <i>b</i> or Ø | <i>e</i> |
| 1 HUM PL | <i>enden</i> | 𒀭... <i>enden</i> | <i>enden</i> | <i>me</i> | <i>enden</i> |
| 2 HUM PL | <i>enzen</i> | <i>y(>e)</i> ... <i>enzen</i> | <i>enzen</i> | — | <i>enzen</i> |
| 3 HUM PL | <i>eš</i> | <i>n</i> ... <i>eš</i> | <i>eš</i> | <i>nne</i> | <i>ene</i> |

Ø = habitual; *ed(a)* = incompletive). In addition, irregular verbs have a different base in incompletive aspect.

As Table 2 indicates, the morphology of finite verbal forms can be much more complex, although a form may be as simple as a prefix, a base, and a subject affix. In addition to the base changing of irregular verbs, finite intransitive forms are marked for aspect with a suffix and transitive forms are marked with the morphology of the core PNG affixes. Setting aside plural PNG affixes, some of which are poorly attested whereas others are circumfixes, Table 3 shows that alignment in completive aspect follows ergative-absolutive principles; in incompletive aspect, it is nominative-accusative in the first and second persons but tripartite in the third person.

Partly on morphological grounds and partly because they have clausal scope, further bound morphemes can be regarded as clitics. These include an enclitic relativizer-complementizer *𒀭* *a* and a set of proclitics that either connect clauses, such as *u* ‘after’, or change mood or polarity (CISL stands for cislocative):

ḥu-mu-na-ab-šum₂-mu
 ḥu=mu-nn-a-b-šum-u
 M=CISL-3.SING-DAT-DO.3.NHUM-give-SUBJ.3.SING
 ‘he must give it to him’

This example illustrates a further characteristic of the script: There is not always a one-to-one correspondence between sound syllable and sign, [nab] being written <na-ab> and [mun] being written simply as <mu>. The transitive subject prefix is an instance of perseverative assimilation from *e* to *u*.

Neither the cohortative (first person) nor the imperative (second person) distinguishes aspect, having instead hybrid forms that combine completive bases with incompletive direct object affixes. Both delete the singular intransitive and transitive subject and can therefore be regarded as following nominative-accusative alignment. The imperative is further irregular in that it is formed not with prefixes but with suffixes.

Resources

The most comprehensive grammar of Sumerian is Attinger (1993: 141–314), although in places it requires familiarity with an earlier, subsequently expanded, publication, Thomsen (2001). No full print dictionary has been published; a web-based dictionary is under development.

Bibliography

- Attinger P (1993). *Éléments de linguistique sumérienne: la construction de du₁/e/di “dire”*. Fribourg, Switzerland/Göttingen: Éditions Universitaires/Vandenhoeck & Ruprecht.
- Thomsen M-L (2001). *The Sumerian language: an introduction to its history and grammatical structure* (3rd edn.). Copenhagen: Akademisk Forlag.

Relevant Websites

- <http://www-etcs1.orient.ox.ac.uk/> – Electronic Text Corpus of Sumerian Literature.
- <http://psd.museum.upenn.edu/> – Sumerian web-based dictionary.

Swahili

L Marten, School of Oriental and African Studies,
London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Swahili is a Bantu language spoken by over 50 million (first- and second-language) speakers in East Africa, including Tanzania and Kenya, where it is a national language, and parts of Somalia, Uganda, Rwanda, Burundi, the Democratic Republic of Congo, and Mozambique. In terms of classification, Swahili belongs to the Sabaki group of the Northeast Coast Bantu languages, and it is part of group G of Guthrie's (1967–1971) referential classification. Like many Bantu languages, Swahili has elaborate noun class and agreement systems, and complex verbal morphology.

Language History

Swahili has been spoken on the East African coast since approximately 800 A.D., after Bantu-speaking people from the Great Lakes region reached the coast. The earliest Swahili speakers, after the language separated from those Sabaki languages most closely related to it, probably settled in northern Kenya on the mouth of the river Tana. Due to the maritime trading of the Swahili, the language became established in Swahili settlements along the coast from Mogadishu in the north to Cap Delgado in the south. Still today, the majority of first-language speakers of Swahili live on the East African coast of Tanzania and Kenya and the adjacent islands. Through continuous contact with Arab traders, many Swahili became Muslims, and a large number of loanwords from Arabic have entered the language over the centuries, leading sometimes to the mistaken belief that Swahili is a mixed language. The Swahili coastal city-states became important centers of the Indian Ocean trade, and in the wake of increasing political and economic power, Swahili poetry flourished, in particular in the Lamu (Kiamu), Pate (Kipate), and Mombasa (Kimvita) dialects, with the earliest surviving Swahili manuscripts, written in Arabic script, dating to the first half of the 18th century. In the 19th century, Zanzibar became part of, and indeed the capital of, the Sultanate of Oman, and the Zanzibar Swahili dialect Kiunguja became more prestigious. During this period, Swahili traders established trade routes into the area beyond the coast, and the language spread with it. During the colonial period, Swahili was

used as a language of administration, especially by the German colonialists in Tanganyika, but it was also a language of interethnic communication in the anticolonial struggle. After independence, Swahili became the national language of Tanzania and Kenya, in both countries sharing the status as official language with English. In Tanzania, and to a lesser extent in Kenya, Swahili is widely used in public administration, education (especially primary education), and the media. Especially in the urban centers, Swahili is increasingly the first language of younger Tanzanians and Kenyans. Swahili is also used to varying degrees in Somalia, Uganda, Rwanda, Burundi, Democratic Republic of Congo, and Mozambique. Outside of East Africa, there are Swahili-speaking communities in the Gulf states and in many Western countries, including the UK (often East-African Indians), and Swahili is taught as a foreign language in language schools and universities throughout the world.

Standard Swahili

There are a number of Swahili dialects spoken today, including the dialects of the Lamu archipelago (Kiamu, Kisiu, Kipate) and Kimvita, associated with classical Swahili literature, and Kiunguja, the dialect of Zanzibar town. Chimwiini and Bajuni, the traditional Swahili dialects of the Somali coast, are currently highly endangered due to the displacement of the Swahili-speaking communities in Somalia, and there are today probably more speakers in Kenya. Comparatively little is known about the more southern Swahili dialects such as Kingome spoken on Mafia island. A distinct variety of Swahili is also spoken in Lubumbashi and the Shaba province in the Democratic Republic of Congo. More recently, distinct urban varieties of Swahili are emerging, for example, the mixed code Sheng of Nairobi. The most important variety of Swahili today is the so-called Standard Swahili (*Kiswahili sanifu*). Since the beginning of the 19th century, various bodies, political and missionary, have made proposals for the development of a standard variety of Swahili. Missionaries began to write Swahili in Roman script, and in 1930 the British-run Inter-territorial Language Committee established the standard form of Swahili based on Kiunguja, which is used today. After independence, strong efforts were made by East African governments to develop Swahili further through research as well as through vocabulary development and standardization, e.g., through the *Baraza la Kiswahili la Taifa* (National Swahili Council) and the *Taasisi*

ya *Uchunguzi wa Kiswahili* (Institute for Swahili Research) in Tanzania. While the development and status of Swahili is often, and rightly, cited as a successful example for the use of an African language as a modern national and official language in post-colonial Africa, it is also leading to an increasing endangerment of Swahili dialects and many of the about 200 languages spoken in Kenya and Tanzania, a problem which has been addressed only recently.

Structure

Swahili exhibits typical Bantu structural characteristics such as an articulated noun class system and morphologically marked agreement between different constituents of clauses and sentences. The morphology is complex and Swahili is often classified as an agglutinating language. Word order, especially within the sentence, is syntactically comparatively free and often motivated by information structure. A remarkable difference from most other Bantu languages is the absence of tone in Swahili.

Noun Classes

A noun class system can be thought of as being halfway between a grammatical gender system as in German or French, and classifier systems as found, for example, in Thai or Chinese. In Swahili, every noun is assigned to a specific noun class, and noun classes are in general marked by a class prefix. Thus, for example, the word *mtoto* 'child' consists morphologically of the noun class prefix *m-* and

the stem *-toto*. Noun classes often express number distinction, so that *watoto*, with a different noun class prefix *wa-*, means 'children.' It is customary in Bantu linguistics to group noun classes according to a numerical system first proposed by Bleek (1869) (see Table 1).

Swahili has 16 different noun classes, which have a more or less transparent semantic base. Nouns in classes 1 and 2 denote only humans (but not all humans are in class 1/2), class 14 is used to refer to abstract qualities, class 15 has verbal infinitives, and classes 16–18 are locative classes. For the remaining classes, the semantic base is less obvious. For example, class 3/4 contains a number of words denoting plants and trees, class 9/10 contains names of animals, and class 6 contains liquids. However, in all of these, there are many words that do not fit a semantic characterization. Another use of the noun classes is for nominal derivation, by shifting nouns from one class to the other. For example, shifting nouns into class 7/8 denotes diminutive: *kitoto* 'a small child,' while class 6 can be used to express a group of individuals, rather than just plurality: *fisi* (class 10), 'hyenas,' *mafisi* (class 6) 'a pack of hyenas.'

Agreement

The noun classes are important for the agreement system of Swahili, as adjectives, demonstratives, and relative clauses show their syntactic relationship with their nominal head through agreement affixes, or concords. Similarly, verbal agreement morphology

Table 1 Swahili noun classes and agreement

| Class | Noun class prefix ^a | Example word | Concord ^b | Relative concord | Possessive concord | Dem prox | Dem ref | Dem non-prox |
|-----------------|--------------------------------|------------------|----------------------|------------------|--------------------|----------|---------|--------------|
| 1 | m | mtu 'person' | a/yu | ye | wa | huyu | huyo | yule |
| 2 | wa | watu 'people' | wa | o | wa | hawa | hao | wale |
| 3 | m | mti 'tree' | u | o | wa | huu | huo | ule |
| 4 | mi | miti 'trees' | i | yo | ya | hii | hiyo | ile |
| 5 | ji | jicho 'eye' | li | lo | la | hili | hilo | lile |
| 6 | ma | macho 'eyes' | ya | yo | ya | haya | hayo | yale |
| 7 | ki | kiti 'chair' | ki | cho | cha | hiki | hicho | kile |
| 8 | vi | viti 'chairs' | vi | vyo | vya | hivi | hivyo | vile |
| 9 | n | ndege 'bird' | i | yo | ya | hii | hiyo | ile |
| 10 | n | ndege 'birds' | zi | zo | za | hizi | hizo | zile |
| 11 | u | ubao 'board' | u | o | wa | huu | huo | ule |
| 14 | u | uhuru 'freedom' | u | o | wa | huu | huo | ule |
| 15 | ku | kuimba 'to sing' | ku | ko | kwa | huku | huko | kule |
| 16 ^c | pa | | pa | po | pa | hapa | hapo | pale |
| 17 ^c | ku | | ku | ko | kwa | huku | huko | kule |
| 18 ^c | mu | | mu | mo | mwa | humu | humo | mule |

^aNoun class prefix is also used for adjective agreement.

^bConcord is used as SM, OM (except in class 1, where SM = *a-*, OM = *m(w)-*).

^cThere are no words in classes 16–18; these are only used in agreement.

marks subjects and objects of the verb. For example, in (1), the demonstrative pronoun and the adjective show their syntactic relation to the head noun *vitabu* (of class 8) by the concord morpheme *vi-*.

- (1) *vi-tabu vi-le vi-zuri*
books those beautiful
 'those beautiful books'

The shape of agreement affixes is not always identical with the noun class prefix. While the agreement morpheme is identical to the noun class prefix in the case of adjective agreement, with demonstratives, it may have a different shape, as for example with a class 6 head noun, where the noun class prefix is *ma-*, but the agreement morpheme of the demonstrative is *ya-* (see **Table 1** for an overview of these forms in all classes):

- (2) *ma-chungwa ya-le ma-zuri*
oranges those beautiful
 'those beautiful oranges'

Verbs show agreement with subjects and objects, by means of subject (SM) and object markers (OM), as in (3):

- (3) *m-toto a-li-wa-angali-a wa-zazi w-ake*
child SM1-PAST-OM2- NP2- CD2-
look_at-FIN parents his/
her
 'the child looked at his/her parents'

The term 'agreement' in relation to verbs can be misleading, as no overt noun phrases are needed for a grammatical sentence; in (4) the subject and object marker function more like pronouns in languages like English:

- (4) *a-li-wa-angali-a*
SM1-PAST-OM2-look_at-FIN
 's/he looked at them'

The subject marker is an obligatory part of the inflected verbs in most tenses, while the object marker is (near) obligatory with human objects, but is used according to semantic and pragmatic considerations (for example, to indicate a specific object or discourse topic) with all other classes.

A number of aspects of the Swahili agreement system pose interesting problems from a theoretical perspective, for example the resolution of agreement with conjoined NPs, or the exact characterization of the status of object agreement. Furthermore, the relation of the Swahili agreement system to the systems of other Bantu (and indeed non-Bantu) languages provides a good testing ground for comparative and historical studies.

Verbal Morphology

Inflected Swahili verbs are, as already seen above, morphologically comparatively complex. In a morphological template for Swahili verbs, ten positions can be identified. Not all of these positions can be filled at the same time, but normally, at least positions 2, 4, 8, and 9 are filled (as in example [5]). Six and seven positions are filled in (6) and (7):

| 1 | 2 | 3 | 4 | 5 | 6 |
|---------|----|---------|--------|----------|--------|
| Pre | SM | Post | Tense | Relative | Stem |
| Initial | | Initial | Marker | Marker | marker |
| Neg | | Neg | | | |
| | 7 | 8 | 9 | 10 | |
| | OM | Verbal | Final | Post | |
| | | Base | | Final | |
| | | | | Plural | |

- (5) *wa-ta-som-a*
SM2-FUT-read-FIN
 'they will read'
- (6) *wa-na-o-ku-j-a*
SM2-PRES-REL2-STEM-come-FIN
 'they who come'
- (7) *ha-wa-ta-ku-ambi-e-ni*
NEG-SM2-FUT-OM2-tell-FIN-PL
 'they will not tell you (pl.)'

In addition to inflectional morphology, verbs can be modified by a number of derivational suffixes, or extensions, suffixed to the verbal root before the final. For example, the causative of *soma* 'read' is *somesha* 'cause to read, teach.' Verbal extensions change the meaning of the base verb and in many cases interact in complex ways with the valency of the base. Among the most productive extensions in Swahili are passive (*-w-*), causative (*-ish-*, *-esh-*), applicative (*-i-*, *-e-*), neutro-passive (*-ik-*, *-ek-*), separative (*-u-*, *-o-*), reciprocal (*-an-*), and stative-positional (*-am-*). The surface forms of these morphemes is determined by phonological processes such as vowel harmony. For example, *funga* 'tie, open,' *fungua* 'untie, close,' *fungia* 'tie for/with someone/something,' *fungika* 'be closable,' *fungana* 'fasten together,' *fungwa* 'be closed,' *funguliwa* 'be opened' (separative and passive), and *fungiana* 'tie for each other' (applicative and reciprocal). The last two examples show that more than one extension can be used. The exact meaning and function of extended verbs depends very much on the meaning of the base verb and on the (syntactic and nonsyntactic) context in which they are used.

Syntax

The basic word order of Swahili in the phrase is head-modifier, and SVOA in the sentence (8). However,

word order can be changed to adapt to the specific discourse-pragmatic situation. Often focused elements are placed at the right periphery of the sentence or phrase, and topicalized elements at the left periphery (9, 10, adapted from Ashton, 1947: 301):

- (8) Asha a-li-m-kut-a Juma njia-ni.
Asha SM1-PAST-OM1- Juma 9.street-LOC
meet-FIN
'Asha met Juma in the street'
- (9) ndoo hizi zi-jaz-e ma-ji
10.buckets these OM10-fill- NP6-water
SUBJ
'these buckets, fill them with water'
- (10) zi-jaz-e ma-ji ndoo
OM10-fill-SUBJ NP6-water 10.buckets
'fill the buckets (not tin cans) with water'

Within the noun phrase, a common strategy to change word order is so-called possessor raising, where in a genitive construction the possessor, which normally follows the possessed, is fronted. In (11), within the subject NP, *mtoto yule* is possessor-raised (cf. *mambo ya(ke) mtoto yule yamenichosha*). In (12), *Sudi* is possessor-raised (cf. *wale wanaojua tabia ya(ke) Sudi*):

- (11) m-toto yule ma-mbo yake
NP1-child this NP6-affairs his
ya-me-ni-chosh-a
SM6-PERF-OM1sg-make_tired-FIN
'as for this child, his affairs make me tired'
- (12) ... hasa wale
... especially those
wa-na-o-m-ju-a Sudi tabia yake
SM2-PRES-REL2-OM1-know-FIN Sudi character his
'... especially those who know Sudi's character'
(Kibao, 1975: 50)

Note that in (12), the object marker agrees with *Sudi*, which is not actually the structural object of *-jua*. Similarly, Maw (1970) reports that in sentences such as (11), 'subject' agreement both with subject, as in (11), but also with the possessor-raised topic are possible (i.e., *amenichosha*). As mentioned above, further work is needed on the analysis of agreement.

Another area of interaction between word order, syntactic function, and agreement are so-called locative inversion structures. In (13), the locative is fronted and the logical subject follows the verb (cf. *watu wengi wamelala humu nyumbani*). Note that the subject marker agrees with the locative phrase, making it the grammatical subject. In (14), the subject marker 'agrees' with an unexpressed locative, and

the subject follows the verb (cf. *hotuba mbali mbali zikatolewa*):

- (13) humu nyumba-ni m-me-lal-a
in_here house-LOC SM18-PERF-sleep-FIN
watu wengi
people many
'many people are asleep in this house'
(Ashton, 1947: 300)
- (14) pa-ka-tol-ew-a hotuba mbali mbali
SM16-CONSEC- speeches different
take_out-PASS-FIN
'and there were held different speeches'
(Kibao, 1975: 50)

Like with many Bantu languages, the formal study of Swahili syntax is only at its beginning, and many constructions, including some of the ones mentioned here, await further analysis.

Bibliography

- Ashton E O (1947). *Swahili grammar* (2nd edn.). Harlow: Longman.
- Bleek W H I (1869). *A comparative grammar of South African languages. Part 1: the concord. Section 1: the noun*. Cape Town, London: J. C. Juta and Trübner & Co.
- Githiora C (2002). 'Sheng: peer language, Swahili dialect or emerging creole?' *Journal of African Cultural Studies* 15, 159–181.
- Guthrie M (1967–1971). *Comparative Bantu* (4 vols). Farnborough: Gregg.
- Johnson F (1939). *A standard Swahili-English dictionary*. Nairobi/Dar es Salaam: Oxford University Press.
- Kibao S A (1975). *Matatu ya thamani*. Nairobi: Heinemann.
- Krifka M (1983). *Zur semantischen und pragmatischen Motivation syntaktischer Regularitäten: Eine Studie zur Wortstellung und Wortstellungsveränderung im Swahili*. Munich: Fink.
- Krifka M (1995). 'Swahili.' In Jacobs J, Stechow A von, Sternefeld W & Vennemann T (eds.) *Syntax: An international handbook of contemporary research*. Berlin, New York: Walter de Gruyter. 1397–1418.
- Marten L (2000). 'Agreement with conjoined noun phrases in Swahili.' *Afrikanistische Arbeitspapiere* 64: *Swahili Forum* VII, 75–96.
- Maw J (1970). 'Some problems in Swahili clause structure.' *African Language Studies* 11, 257–71.
- Miehe G & Möhlig W J G (eds.) (1995). *Swahili-Handbuch*. Cologne: Köppe.
- Mohamed M A (2001). *Modern Swahili grammar*. Nairobi: East African Educational Publishers.
- Nurse D & Hinnebusch T J (1993). *Swahili and Sabaki: a linguistic history*. Berkeley: University of California Press.
- Nurse D & Spear T (1985). *The Swahili: reconstructing the history and language of an African society, 800–1500*. Philadelphia: University of Pennsylvania Press.
- Polomé E C (1967). *Swahili language handbook*. Washington DC: Center for Applied Linguistics.

- Sacleux C (1909). *Grammaire swahilie*. Paris: Procure de PP. du Saint-Esprit.
- Sacleux C (1939). *Dictionnaire swahili-français*. Paris: Institut d'Ethnologie.
- Schadeberg T C (1992). *A sketch of Swahili morphology*. Cologne: Köppe.

- Taasisi ya Uchunguzi wa Kiswahili (1981). *Kamusi ya Kiswahili Sanifu (Dictionary of Standard Swahili)*. Nairobi, Dar es Salaam: Oxford University Press.
- Whiteley W H (1969). *Swahili: the rise of a national language*. London: Methuen.

Swedish

K Börjars, The University of Manchester, Manchester, UK

© 2006 Elsevier Ltd. All rights reserved.

Swedish is spoken natively by ~8.5–9 million people in Sweden and by ~250 000–300 000 people in Finland. It is a Germanic language, part of the North Germanic branch, along with Norwegian, Danish, Icelandic, and Faroese. There is a fair degree of mutual intelligibility between Swedish and the other so-called mainland Scandinavian languages. The early historical stages of the North Germanic languages are normally divided into an eastern group (the dialects of the present Norway, Iceland, and Faroe Islands) and a western group (the dialects of Sweden and Denmark). As the language situation developed and Danish and Swedish crystallized as two separate languages, a north–south division became increasingly appropriate. The year 1526 is normally taken as the beginning of Modern Swedish, at which time Sweden won independence from Denmark and the first Swedish translation of the New Testament began to be circulated.

The language was originally written in runes carved in stone, but early Christian missionaries brought the idea of writing on parchment and with it the Latin alphabet. There are books preserved from as early as the beginning of the 13th century. The modern Swedish alphabet contains 28 letters; the same as those of the English alphabet, except that there is no 'w' and there are three additional letters at the end of the alphabet (å, ä, and ö). Swedish underwent a spelling reform in 1906 and its spelling is now quite regular, though there are a few sounds that can be represented in a number of different ways, most notoriously /f/, which can be spelled *sk*, *sj*, *stj*, *skj*, *ch*, and *sch*, and /j/, which can be spelled *j*, *gj*, *hj*, *dj*, and *lj*.

Phonologically, Swedish is characterized by a relatively large number of vowels. The 18 vowels are frequently grouped into nine pairs; the main difference within each pair is length, but there are also

associated differences in quality. Standard Swedish does not have phonemic diphthongs, though the pronunciation of long vowels may involve some diphthongization. The consonants are also realized as long or short, with little or no difference in quality. Long sounds may only occur in stressed syllables and every stressed syllable must contain either a long vowel or a long (or double) consonant. When any of the consonants /t d s n l/ immediately precede /r/, the two consonants are then realized as a retroflex, /ṭ ḍ ṣ ṇ ḷ/, respectively.

Swedish, apart from the Swedish spoken in Finland, makes a distinction that is often referred to as tonal, i.e., there is a difference between Accent I (or *akut accent*) and Accent II (or *grav accent*). The distinction is one of word accent. The difference between the two accents is mainly one of pitch, but Accent II, which is limited to bi- and polysyllabic words, has an effect of some secondary stress on the syllable immediately following the syllable with main stress. There are a number of minimal pairs in the language, distinguished only by Accent I versus Accent II, as in Examples (1a) and (1b), where I or II indicates the type of accent (abbreviation: DEF, definite):

- (1a) 'tomten "tomten
yard.DEF gnome/father Christmas.DEF
- (1b) 'anden "anden
duck.DEF spirit.DEF

Swedish verb morphology is relatively simple, with no agreement marking in any tense. The present–past distinction is made morphologically, whereas perfect aspect is marked by the auxiliary *ha* 'have,' followed by a form of the verb referred to as the 'supine.' For passive, there is both a morphological and a syntactic version, and a number of subtle factors influence the choice between the two. A paradigm for the verb *kittla* 'tickle.INFINITIVE' is provided in Example (2) (SG, singular; PL, plural; PRES, present; PERF, perfect; S PASS, BLI PASS, morphological and syntactic passive; FEM, feminine):

| (2) | SINGULAR | PLURAL | PRESENT | PAST | PERFECT (PRES) | -S PASSIVE | BLI PASSIVE |
|-----|----------|--------|---------|----------|----------------|------------|--------------|
| 1 | jag | vi | kittlar | kittlade | har kittlat | kittlas | blev kittlad |
| 2 | du | ni | | | | | |
| 3 | hon(FEM) | de | | | | | |

Noun phrases, on the other hand, have richer morphology, including agreement marking on modifiers. The masculine and feminine genders have merged into one, usually referred to as common gender (or *utrum*) and marked by *-n* in singular, which contrasts with neuter, marked by *-t*. In the plural and definite noun phrases, the gender distinction is neutralized. Definiteness is marked morphologically on nouns, and a singular count noun in its definite form can function as a referential noun phrase without any need for a syntactic determiner (see Examples (3a)–(3c)); without the definiteness marking, a syntactic determiner is required for the noun to function as a full noun phrase, as illustrated by Examples (4a) and (4b). When a modifier precedes the noun, a syntactic determiner is also required. In most cases, the noun retains its morphological marking, giving rise to so-called double definiteness, as Examples (5a)–(5c) show. Most modifiers show agreement with respect to gender (in singular indefinite), number, and definiteness. This is illustrated for definite noun phrases (Example (5)) and for indefinite ones (Example (6)). The definite–indefinite distinction on modifiers is commonly referred to as a weak–strong distinction in the literature. Case marking is found only on pronouns in Swedish (DEF, definite; COM, common; NEUT, neuter; INDEF, indefinite).

(3a) gris-en
pig-DEF.COM.
'the pig'

(3b) djur-et
animal-DEF.NEUT
'the animal'

(3c) gris-ar-na
pig-PL-DEF.PL
'the pigs'

(4a) en gris
a.COM pig
'a pig'

(4b) ett djur
a.NEUT animal
'an animal'

(5a) den ren-a gris-en
the.COM clean-DEF pig-DEF.COM
'the clean pig'

(5b) det hungrig-a djur-et
the.NEUT hungry-DEF animal-DEF
'the hungry animal'

(5c) de ren-a gris-ar-na
the.PL clean.DEF pig-PL-DEF.PL
'the clean pigs'

(6a) en ren gris
a.COM clean.COM.SG.INDEF pig
'a clean pig'

(6b) ett hungrig-t djur
a.NT hungry-NEUT.SG.INDEF animal
'a hungry animal'

(6c) två hungriga grisar/ djur
two hungry.PL pigs/ animals
'two hungry pigs/animals'

There is also a participle form, distinct from the supine, that is used attributively and predicatively and that agrees in gender and number, in a way similar to adjectives; this is illustrated in Examples (7a) and (7b) (PART, participle):

(7a) Brevet är skrivet
letter.DEF.NEUT be.PRES write.PART.NEUT
för hand.
by hand.
'The letter is written by hand.'

(7b) ett slarvigt skrivet
a.NEUT carelessly write.PART.NEUT.SG
brev
letter
'a carelessly written letter'

A striking property of Swedish is also that the possessive determiner exists in a reflexive and a non-reflexive form. The reflexive is used roughly in those environments in which a pronoun replacing the whole noun phrase would have to occur in its reflexive form. In Example (8a), then, *Björn* is eating someone else's sandwiches, whereas in Example (8b), he is eating his own sandwiches (POSS, possessive; MASC, masculine; REFL, reflexive):

(8a) Björn_i äter hans_j smörgåsar. [i ≠ j]
Björn eat.FIN POSS.MASC sandwich.PL
'Björn is eating his sandwiches.'

(8b) Björn_i äter sina_i smörgåsar.
Björn eat.FIN POSS.REFL.PL sandwich.PL
'Björn is eating his (own) sandwiches.'

The possessive reflexive agrees with its noun for number and gender much like an adjective; it does not, however, mark the gender of the possessor, unlike the nonreflexive form.

Like the other North Germanic languages, Swedish is a verb-second language. This means that main clause word order is built around the finite verb in the second position. The initial phrase will either be the subject or will have some special information structural status, such as topic or focus, or will be an adverbial phrase, often a so-called scene-setting adverbial phrase. Some examples are provided in the following sentences:

- (9a) Philip gillar matematik.
Philip like.PRES mathematics
'Philip likes mathematics.'
- (9b) Dinosaurier gillar Nils.
Dinosaur.PL like.PRES Nils
'Nils likes dinosaurs.'
- (9c) Under sängen hittade Ellen
under bed.DEF find.PAST Ellen
inte några sockor.
not some sock.PL
'Ellen didn't find any socks under the bed.'

Any phrasal constituent can then precede the finite verb, including clauses, as illustrated in Example (10):

- (10) Att han måste ha hjälm när
that he must have.INF helmet when
han cyklar gillar Robin inte.
he cycle.FIN like.FIN Robin not
'Robin does not like the fact that he has
to wear a helmet when he cycles.'

The word order in the part of a main clause that follows the finite verb is usually described as relatively firm, with the order shown in Example (11):

- (11) Main clause word order:
INITIAL CONSTITUENT → FINITE VERB → SUBJECT →
ADVERBIAL → NEGATION → NONFINITE VERBS →
OBJECTS/COMPLEMENTS

There is, however, some variation in word order also in this part of the sentence, motivated by factors such as information structure and scope. For instance, the subject *Robin* and the negation in Example (10) could change places.

The word order in subordinate clauses differs from that in main clauses in that the verb does not normally occur in second position, and it can only do so under certain very specific circumstances. Instead, the finite verb follows the subject and adverbials – in particular,

the negation. The subordinate version of Example (9c) would then be as in Example (12):

- (12) ...att Ellen inte hittade några
that Ellen not find.PAST some
sockor under sängen.
sock.PL under bed.DEF
'... that Ellen didn't find any socks under the bed.'

Naturally, there are many Swedish dialects, two of which deserve mention here. The first is the Swedish spoken natively in Finland: this dialect is quite distinct from the Swedish spoken in Sweden in phonology, lexicon, and syntax, one of the most striking differences being the lack of the two tones previously described (*akut accent* and *grav accent*). The other variety of Swedish of note is spoken in a small area roughly in the middle of Sweden; this dialect, *Älvdalen*, is closer to older forms of Swedish in that it preserves more morphological marking (for instance, case marking and agreement on the finite verb).

Sweden has long had a generous immigration policy and hence speakers of a large number of languages now live in Sweden. Though it is a controversial issue, there have been claims that a new variety of Swedish is emerging, namely, that spoken natively by children born in Sweden to parents who are not native speakers of Swedish. In the literature, a number of terms have been used to refer to this variety of the Swedish language, the most neutral being *Svenska på mångspråkig grund* 'Swedish on a multilingual basis.'

In 1786, *Svenska Akademien* 'The Swedish Academy' was set up to promote the purity of the language. The Academy continues to be responsible for publishing the major monolingual Swedish dictionary, *Svenska akademiens ordbok*, available online at www.saob.se. The Academy has also published a four-volume grammar of Swedish (see Teleman *et al.*, 1999). An excellent collection of corpora of Swedish, written and spoken, modern and historical, is publicly available at *Språkbanken* 'the language bank' at Gothenburg University (spraakbanken.gu.se).

Bibliography

- Holmes P & Hinchliffe I (1994). *Swedish: a comprehensive grammar*. London: Routledge.
- Teleman U, Hellberg S & Andersson E (1999). *Svenska akademiens grammatik*. Stockholm: Nordstedts.

Syriac

G Kiraz, Beth Mardutho: The Syriac Institute,
Piscataway, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Syriac is a form of Aramaic, a Semitic language whose many dialects have been in continuous use since the 11th century B.C. Syriac is by far the most attested dialect of Aramaic. It is used today in two forms: Classical Syriac, which is a literary form of the language, and Vernacular Neo-Syriac, which consists of many regional dialects. Syriac is used by Christian communities in the Middle East, known as Syriacs, Assyrians, Chaldeans, and Maronites; and in the Indian state of Kerala, primarily as a liturgical language, by communities known as the St Thomas Christians. There are today only a few hundred speakers of Classical Syriac, fewer than one million speakers of Vernacular Neo-Syriac, but over 10 million who consider Classical Syriac their liturgical language. Classical Syriac exists in two main dialects, West Syriac and East Syriac, the difference between them being minor phonological variations.

Historically, the earliest dated Syriac inscription is from 6 A.D., and the earliest parchment, a deed of sale, is from 243. The earliest dated manuscript was produced in November 411, probably the earliest dated manuscript in any language. Within a few centuries from its origin, Syriac produced a wealth of literature that surpassed all other Aramaic dialects. Early literature was produced in Mesopotamia, especially in and around Edessa, by pagans, agnostics, Jews, and Christians. The literature of the first three centuries consists mostly of anonymous texts whose date and origin cannot be established. The 4th century witnessed the first major writings that survive to this day. The 5th to 9th centuries mark the Golden Age of Syriac, with more than 70 important known authors, not counting numerous anonymous works and lesser authors. These writings cover philosophy, logic, medicine, mathematics, astronomy, alchemy, history, theology, linguistics, and literature. Under the Arabs, Syriac was the vehicle by which the Greek sciences passed to the Muslim world, and later to Europe through Spain, marking Syriac as an important stage in the history of world civilization. As Arabic began to replace Syriac as the primary language of the Middle East, Syriac became less prominent but has continued to be used until today.

The Syriac writing system makes use of three scripts. The oldest, known as Estrangelo 'rounded,' was fully developed by the 5th century. Later, two

geographic scripts derived from it: West Syriac, whose proper name is Serto, and East Syriac. Early Syriac writing consists of consonants and long vowels only. In the 7th century, a vocalization system was developed and lent itself to Hebrew and Arabic. At the time of Genghis Khan (12th century), the Mongolian script was derived from Syriac.

The phonology of Syriac makes use of 22 consonants, three of which are *matres lectionis* (glottal stop, *u*, and *y*), and seven vowels (five in the case of West Syriac). Six consonants, known by the mnemonic *bgdkpt*, undergo spirantization, where the plosives become fricatives. Traditionally, stress has been assigned to the penultimate syllable in West Syriac, and the final syllable in East Syriac. Syllabification employs long open (CVV) and closed (CVC) syllables. The short vowel of a CV syllable is almost always deleted.

The morphology of Syriac is based on root-and-pattern morphology, in addition to suffixation, prefixation, and circumfixation. Most roots consist of three consonants, although two- and four-consonantal roots exist. Roots that do not contain any of the *matres lectionis* are called 'strong,' and those containing *matres lectionis* are called 'weak' and for the most part undergo various phonological processes. Most words are derived according to a CV template and a vocalism.

Verbs exist in two tenses: perfect, denoting past tense, marked by zero or one suffix; and imperfect, denoting future tense, marked by a circumfix. The imperative is marked by the suffix part of the imperfect circumfix. Closely related to verbs are the participles and the infinitive. Verbal affixes mark number (singular, plural), person (1st, 2nd, 3rd), and gender (masculine, feminine).

Nouns exist in three states: 'absolute' is the basic form and in early Syriac used to indicate nondetermination; 'emphatic,' by far the most frequent, is marked by a gender-sensitive suffix and is used to mark determination; and 'construct' (joining two nouns) is used primarily to mark a genitive like relation. Adjectival forms are formed mostly by one or more suffixes, those with fewer suffixes belonging to earlier periods of the language. More complex nouns are formed by formative prefixes, and may also contain suffixes.

Personal pronouns either stand on their own, or are in the form of suffixes; they are also either in subject form or object form. Demonstrative and interrogative pronouns stand alone. The relative pronoun is in the form of a prefix.

The sentence structure does not put hard constraints on word or clause order, though idiomatic construction is not very free. Nominal sentences have a noun, an adjective, or an adverbial expression as a predicate. Copulative sentences are joined together with a conjunction in the form of a prefix (in the case of 'and') or a stand-alone word (in the case of 'or'). Syriac also uses relative clauses, marked by the prefix *d*, indirect interrogative clauses, marked by a particle, and conditional clauses, also marked by a particle.

The Syriac lexicon is either arranged by root, or in a quasialphabetical order (in the latter case, derivations of the verb with prefixes appear in the unprefixated 3rd singular masculine form). The primary Syriac lexica in use were all composed in the 19th and early 20th centuries.

Bibliography

- Brock S (2001). *The hidden pearl: the Aramaic heritage*, 3 vols and videos. Piscataway, NJ: Gorgias Press; Italy: Trans World Film.
- Healy J (2005). *Leshono Suryoyo: first studies in Syriac*. Piscataway, NJ: Gorgias Press.
- Kiraz G (2005). *The Syriac primer*. London: T. & T. Clark International.
- Nöldeke T (2001). *Compendious Syriac grammar*. Crichton J A (trans.). Winona Lake, IN: Eisenbrauns.
- Payne Smith R (1998). *A compendious Syriac dictionary: founded upon the Thesaurus Syriacus of R. Payne Smith*. Payne Smith J (ed.). Winona Lake, IN: Eisenbrauns.
- Robinson T H & Coakley J F (2003). *Robinson's paradigms and exercises in Syriac grammar* (5th edn.). Oxford: Oxford University Press.
- Thackston W M (1999). *Introduction to Syriac*. Bethesda, MD: IBEX.

T

Tagalog

J U Wolff, Cornell University, Ithaca, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Tagalog, spoken in the Philippines, is a member of the Austronesian group of languages. The Austronesian languages are descended from Proto-Austronesian, which is believed to have developed on the Asian mainland and to have been brought to Taiwan by around 6000 B.C., whence its descendents spread through the Philippines and Indonesia eastward to the islands of the Pacific. The earliest documents in Tagalog date from a few decades after the first Spanish colonization in 1564. The pre-Hispanic Tagalogs had a syllabary called *Alibata*, which has been recorded, but if there was any written literature, none of it survives. The end of the 16th century and the beginning of the 17th saw the publication of a catechism *Doctrina Cristiana*; a magnificent and thoroughgoing dictionary, a grammatical description, and a textbook that purports to teach Spanish to Tagalog speakers. These provide good documentation of what Tagalog was like at the time, and indeed the language of these texts is readily understandable today. Other literature, mostly poetry, dates from the middle of the 19th century. It was only in the beginning of the 20th century that prose literature and other types of writing were published in Tagalog. Education in the medium of Tagalog was not introduced until the 1960s, and to this day, English predominates as the medium of instruction at all levels.

Tagalog has a unique status among the more than 100 indigenous languages of the Philippines in that it is the national language alongside of English. Although English still predominates in the Philippines as the language of education, public affairs, and formal occasions, Tagalog is increasingly coming into use in these settings, particularly in those areas in which Tagalog is spoken natively. At the time of the Philippine Commonwealth, Tagalog was spoken by less than a quarter of the population of the Philippines. To avoid political controversy among the speakers of other languages, the fiction was adopted that this language, with some modification of vocabulary taken

from other major Philippine languages, was an amalgam of these languages. As such it was called 'Pilipino.' In the 1970s, a new fiction was adopted, that the amalgamation of Philippine languages that was to serve as the national language was composed of a larger number of the indigenous languages than Pilipino had been, and this new language was termed 'Filipino.' However, the terms 'Pilipino,' 'Filipino,' or 'Tagalog' all refer to one and the same language, and all three terms are commonly used to refer to it.

Tagalog is spoken indigenously in the Manila region and in the provinces surrounding it. As such, it is the language associated with the seat of Philippine power and culture, and has acquired a special cachet or prestige. In the last few decades, Tagalog has spread far beyond its original home to urban areas to the south and the north, especially those that have seen a large influx of immigrants from other regions, although in Mindanao there is strong competition from Cebuano, and in the north competition from Ilocano. However, at this point, Tagalog has become the native language of approximately one-third of the population of the Philippines and is ever increasing in number of speakers. In addition, Tagalog press, TV, and cinema, and most importantly, population mobility, have spread the knowledge of Tagalog throughout the nation, so that only few people, and those mostly in the oldest generation, do not have at least a passive knowledge of Tagalog. Further, although loyalty to the native language is strong in the Philippines (few of the indigenous languages are in danger of dying, even though some have small numbers of speakers), it is becoming increasingly acceptable to use Tagalog in social settings (even in non-Tagalog regions), where family members and guests use Tagalog instead of the native language. This usage usually occurs on the part of native sons, who have moved to Tagalog-speaking regions for employment, or their children. Many have become dominant in Tagalog. Speaking in Tagalog in a group where everyone else is speaking the native language is not remarked upon, and in fact, there is a certain prestige attached to speakers who do this, as it is a sign of having made good in the outside world. Abroad, Tagalog has become the mark of Philippine

national identity and is used by Filipinos with other Filipinos, no matter what region they come from, and internally, Tagalog is well on its way to becoming the lingua franca of a multilingual nation.

What Tagalog Is Like

Tagalog has a simple phonology. The consonants, vowels, and diphthongs are as follows (Table 1). Note that there are long and short vowels: the long vowels are marked with an accent.

Glottal stops in standard Tagalog occur only before a pause. If a word with a glottal stop at the end of it occurs in a phrase with another word following it, the glottal stop is lost, and there is compensatory lengthening of the preceding vowel:

Wala? ‘not’ + *na* ‘any longer’ produces *walá na* ‘no longer’

The spelling system ignores important parts of the phonology and does not recognize long versus short, although pedagogical texts use a cumbersome system to indicate these partially. The system also does not indicate /ʔ/. The palatals /c/, /j/, and /ʃ/ are written *ts*, *dy*, and *sy* respectively. The phoneme /ŋ/ is written *ng*:

/cinílas/ *tsinilas* ‘slippers’; /jip/ *dyip* ‘jeep’; /ʃa ŋ á pala / *syá nga pala* ‘by the way’

At the time the Spaniards first came to the Philippines, Tagalog clearly did not have this phonological system. The phonemes in parentheses in Table 1 show sounds that have been added to Tagalog since that time. This addition is proven not only by comparing Tagalog with other Philippine languages (i.e., doing historical reconstruction) but also by the treatment of loanwords from Spanish at the early time and the modern time. An example is the Spanish word for ‘hat’ *sombrero*, which was borrowed twice in Tagalog: once early on and then again later. These two words are now perceived to be two different lexical items and refer to different things, but their phonemic make-up show how Tagalog has expanded its phonology: it has added /o/ and /e/, it has come to allow a consonant clusters with /r/, and has come to allow a vowel other than /a/ to occur three or more syllables from the end:

Table 1 Tagalog phonology

| Consonants | | | | Vowels and Diphthongs | | | |
|------------|--------|-----|---|-----------------------|--------|--------|------|
| p | t | (c) | k | ʔ | i, í | | u, ú |
| b | d | (j) | g | h | (e, é) | (o, ó) | |
| m | n | | ŋ | | | a, á | |
| | s, ʃ | | | | iw | | uy |
| w | l, (r) | y | | | | ay, aw | |

sambalílu?, ‘conical sun hat of the native variety’ *sombrero* ‘western hat’

Not all of this is due directly to Spanish influence. Much has to do with internal developments in Tagalog itself, but the contact with Spanish was a catalyst or facilitated some of these developments, in that Spanish words pronounced closer to the Spanish pronunciation made certain rare combinations or types more common.

In grammar, Tagalog is characterized as a synthetic rather than an analytic language (English is an example of an analytic language). In Tagalog, single words containing a root plus affixes of all sorts express what in analytic languages would be expressed by a phrase. For example, the single word *pápa-pagparikitin* ‘will cause him or her to make a fire’ expresses what in English takes seven words to express. The root here is *dikit* (the initial /d/ is changed to /t/ by rule that says /d/ between vowels is often changed to /r/).

The verbal system in Tagalog expresses the relation between the verb and a word it refers to: the word referred to may signify the agent, the place, the beneficiary, the instrument, the patient, the thing moved, or the indirect object (depending on the affix). The verb contains what in English would be a verb and a preposition. An example is the root *pútol* ‘cut’:

- (1) (agent) Ako ang púpútol
I the-one-who will-cut
nang táli?
object-maker string
‘Let me be the one to cut the string.’
- (2) (patient) Putúlin mo ang táli?
Cut-it by-you the string
‘Cut the string.’
- (3) (local) Putúlan mo nang
Cut-from-it by-you object-marker
kónti ang kék.
little the cake
‘Cut a little from the cake.’
- (4) (benefactive) Ipútol mo
Cut-for by-you I
ako nang kék.
object-marker cake
‘Cut the cake for me.’
- (5) (instrumental) Itong kutsilyo ang
This knife the-one-that cut-with-it
ipangpútol mo
by-you on string
‘Cut the string with this knife.’

Cutting across this system of voice or prepositional-like affixes is a system of four-way inflections that expresses time (past or present as opposed to future),

ongoing or iterative action, as opposed to a single action, and an imperative inflection, which is also used to express dependence, optativity or uncertainty. For example, sentence (1) above exemplifies future tense, (6a) below exemplifies noncompleted action, (6b), past action, and (6c), uncertain action:

- (6a) Ako ang laging
I the-one-who always cut
pumuputol nang tali?
object-maker string
'I am the one who is in charge of cutting
(literally, always cuts) the string.'
- (6b) Sino ang
Who the-one-who did-cut
pumutol nang tali?
object-maker string
'Who (purposely) cut the string?'
- (6c) Baká pumutol siya nang tali?
lest cut he object- marker string
'He might just cut the string.'

Perfective action is not expressed by verbal inflections but rather analytically (by a phrase).

- (6d) Matagal na akong
long-ago has-been I
pumutol nang tali?
cut object-maker string
'It has been some time since I (purposely) cut the string.'
- (6e) Alas sayis na ako
o'clock six have-done I
puputol nang tali?
will-cut object-maker string
'At six o'clock I will have cut the string.'

There is a large number of derivational affixes that interact with the above-mentioned inflectional affixes to produce verbal forms with a wide number of meanings. Some of these affixes are applicable to almost all roots, some are more limited in their distribution. The most productive are the causative affix *pa-* and the potential affix *ka-*, both of which are addable to almost all roots that take verbal affixes. More than one derivational affix may occur within a verb. There are affixes that transitive verbs, others that form verbs of reflexive action, those that form plurals, those that indicate an action done by two together, by more than two together, actions done as a favor, actions involving another, actions done by accident, and so forth. There is also a large number of adjective and nominal derivations.

Here are a few examples from the root *sáma*, a small percentage of the total number of derivational forms that occur with this root:

No derivational affix, 'go along':

- (7) Ayó kong sumáma
not-want I go-along
'I don't want to go along.'

With *pag-*, transitivizer: 'take something. along somewhere':

- (8) Magsáma [=um-+pag + sama] ka nang
Bring-along you object-marker
tá'o pagpunta mo do' on.
person when-go you there
'Bring someone with you when you go there.'

With *pa-*, causative:

- (9) Hindi mo siyang dápat
Not by-you him should
pasamáhin.
cause-him-to-accompany
'You should not allow him to come along.'

With *ka-*, potential action:

- (10) Hindi ka makákasáma [=future
active + potential + sama]
Not you will-be-able-to-go-along
kung iiyak ka.
if cry you
'You won't be able to come along if you are going to cry.'

With *kápa-*, accidental and causative action:

- (11) Nápasáma [= past-passive + ká- yung
accidental action + pa-
causative + sama]
Was-accidentally-caused-to-go-along that
papel sa dala-dala ko
paper with thing-brought my
'I accidentally took that piece of paper together
with the thing I was bringing (That piece of
paper got caught up with the things I was
bringing).'

With *pag-*, 'do together':

- (12) San Miguel, ang bir na may
San Miguel the beer that there-is
pinagsamáhan [= past + local-passive + pag-
+ sama].
be-companions-over-it
'San Miguel, the beer people have
companionship over (while drinking).'

Influences on Tagalog and Tagalog's Influence on Other Languages

Tagalog, as a language of wider communication and as a spreading language, is being simplified. Simplification is most marked in urban areas, and the process

is gradually spreading to the provinces. It undoubtedly begins from errors made by people (immigrants from non-Tagalog regions or Filipino Chinese) who learn Tagalog as a second language and are imitated by native speakers. One simplification is the loss of contrast between long and short vowels in certain syllables, leading to the loss of contrast between the accidental and the potential conjugations, e.g., *nápa-sáma* of example (11) above is pronounced /napasáma/ (which in conservative Tagalog has no meaning). Similarly, *mábibili* 'someone might buy it' (the accidental passive future) is pronounced /mabibili/. Thus, the contrast is lost between *mábibili* 'someone might buy it' and *mabibili* 'is able to buy it'. Another aspect of this simplification is that there is a tendency to drop many of the productive derivations, which are very much alive in conservative Tagalog and exclude any vocabulary but that of the highest frequency. This tendency is exacerbated by the secondary role of Tagalog *vis-à-vis* English in public life and in education.

Mutatis mutandis, Tagalog influences the other indigenous languages of the Philippines. These other languages are replete with Tagalog loanwords that stem from the language's widespread use in the media. In some areas, Tagalog has a more intimate effect. In Samar, for example, where a large portion of the population has work experience in the Manila area and Manila has an especial cachet, the regional language is spoken by many younger people with a clearly observable Tagalog intonation. In the urban parts of the Cebuano speech area, where the managerial class is largely composed of immigrants from Tagalog regions who have learned Cebuano as a second language, complexities of Cebuano syntax that have no analogue in Tagalog are lost or regularized, and this syntax has spread to the younger generation of Cebuano natives who have no Tagalog connection (see **Cebuano**).

Bibliography

- Bautista M L S (1980). *The Filipino bilingual's competence: a model based on an analysis of Tagalog-English code switching*. Canberra: Dept. of Linguistics, Research School of Pacific Studies, Australian National University.
- Belwood P (1991). 'The Austronesian dispersal and the origin of languages.' *Scientific American*, July, 88–93.
- Bloomfield L (1917). 'Tagalog texts with grammatical analysis.' *Studies in Language and Literature* 3(2), 2–4.
- English L J (1986). *Tagalog English dictionary*. Manila: Congregation of the Most Holy Redeemer.
- Gonzalez A (2002). 'Language planning and intellectualization.' *Current Issues in Language Planning* 3(1), 5–27.
- Gonzalez A & Postrado L (1976–1977). 'Dissemination of Pilipino.' *Philippine Journal of Linguistics* 1(2), 60–84.
- Kelz H P (1981). 'Sprachplanung auf den Philippinen und die Entwicklung einer philippinischen Nationalprache.' *Language Problems and Language Planning* 5(2), 115–136.
- San Buenaventura Fr P de (1613). *Vocabulario de lengua Tagala*, printed by Tomás Pinpin and Doimingo Loag; Facsimile copy by Paris-Valencia, Pelayo 7, 46007 Valencia, Spain.
- Schachter P & Otnes F T (1972). *Tagalog reference grammar*. Berkeley: University of California Press.
- Smolicz J J & Nical I (1997). 'Exporting the European idea of a national language: some educational implications of the use of English and indigenous languages in the Philippines.' *International Review of Education* 43(5–6), 507–526.
- Wolff J U (1993). 'Why roots add the affixes with which they occur: a study of Tagalog and Indonesian adjective formations.' In Reesink G P (ed.) *Topics in descriptive Austronesian linguistics* 11. Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië Rijksuniversiteit te Leiden.
- Wolff J, Rao D H & Centeno T (1991). *Pilipino through self-instruction*. Ithaca: Southeast Asia Program, Cornell University.

Tahitian

P Geraghty, University of the South Pacific, Suva, Fiji

© 2006 Elsevier Ltd. All rights reserved.

Tahitian belongs to the Eastern Polynesian branch of the Oceanic subgroup of the Austronesian language family. Its nearest relatives are other Central Eastern Polynesian languages, such as Tuamotuan, Marquesan, and Cook Islands Maori.

Until the early 19th century, Tahitian was spoken by the entire population of the Society Islands, and it remained the main language for most of that century. Annexation by France (in 1880) and educational and social policies have contributed to the decline of Tahitian, particularly in the capital, Pape'ete. At the same time, Tahitian has become the lingua franca of the Marquesas, Tuamotus, Austral Islands, and other parts of French Polynesia, at the expense of their respective indigenous languages, and it is now

estimated to have 150 000 speakers, including some 5000 residents of New Caledonia.

Tahitian has been an official language of French Polynesia, along with French, since 1978, but more *de jure* than *de facto*. It is taught in a small way up to university level. There is a substantial amount of Tahitian language radio and television programming, but no newspaper. The recent (2004) election of a government committed to more independence from France may bring about changes in the use and status of Tahitian.

Tahitian had no traditional written form and was first recorded by 18th-century explorers such as Bougainville and Cook. The latter was responsible for introducing into English the loanwords *taboo* and *tattoo*, from Tahitian *tapu* and *tatau*. A Roman-based alphabet was devised by English-speaking missionaries in 1815 and has remained in use relatively unchanged. Reliable reference works are currently available only in French.

An unusual feature of Tahitian was the custom of 'pi'i,' by which everyday words that constituted parts of chiefs' names were considered taboo, and in many cases the change became permanent. Tahitian is also unique among Pacific languages in having an academy (Fare Vana'a), founded in 1974, that aims to standardize, develop, and promote the language.

The phoneme inventory of Tahitian consists of nine consonants (f, h, m, n, p, r, t, v, and glottal stop) and

10 vowels (a, e, i, o, u, ā, ē, ī, ō, ū). There are no consonant clusters, and syllables are open. In writing, vowel length and glottal stop have often not been marked systematically. Some modern writers and publishers use a macron to indicate a long vowel and an apostrophe to indicate the glottal stop, as recommended by the Fare Vana'a.

There is very little morphophonemics, and most grammatical functions are performed by affixation or the use of pre- and postposed particles. Pronouns distinguish four persons (including first-person inclusive and exclusive) and three numbers (singular, dual, and plural). There are two categories of possession, depending largely on whether or not the possessor has control over the fact of possession. In noun phrases, the order is head + attribute. The basic word order is VSO:

| | | | | | | |
|------------------------|-------|------|-----|-----|------|------|
| 'ua | tai'o | 'oia | i | te | puta | rahi |
| ASP | read | he | OBJ | the | book | big |
| 'he read the big book' | | | | | | |

Bibliography

- Académie tahitienne (1999). *Dictionnaire tahitien-français*. [Papeete:] Fare Vana'a [Académie tahitienne].
- Anonymous (n.d. [1986]). *La grammaire de la langue tahitienne*. [Papeete:] Fare Vana'a.
- Lemaitre Y (1973). *Lexique du Tahitien contemporain*. Paris: ORSTOM.

Tai Languages

D A Smyth, SOAS, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Geographical Location and Number of Speakers

Tai languages are spoken by over 70 000 000 people across a wide area of Asia that extends from Vietnam in the east to India in the west. The most important member of the family is Thai, the national language of Thailand, which accounts for approximately two-thirds of all Tai speakers. The second highest national concentration is in China, where there are an estimated 15 000 000 speakers, mainly in the southwest. Smaller Tai-speaking populations live in northern Vietnam, Laos, Burma, and northern India.

The Tai language family comprises three branches: southwestern, central, and northern. The southwestern group extends over the widest geographical area

and includes the national languages of Thailand and Laos, plus Shan and Khün (spoken in northern Burma); Lü (China-Burma border); Khamti (Burma-India border); and Black Tai (Tai Dam), White Tai (Tai Dón), and Red Tai (Tai Daeng) (Laos-Vietnam border). The central and northern branches are geographically more homogeneous, languages from both groups being spoken in both northern Vietnam and southern China. Central Tai includes Tho (Tày), Longzhou, and Nung, while Northern Tai includes Wu-ming (Northern Zhuang), Yoi (Dioi), and the Bouyei (Pu-yi) languages of China.

Wider Affiliations

Certain lexical and grammatical similarities between the Tai and Chinese languages led linguists in the 19th century to assume that the two groups were related, and until the 1940s this was the widely accepted view.

Since then, however, most authorities have come to believe that there is no such genetic link and that any similarities are due to borrowings. The wider affiliation of Tai languages has been the subject of considerable scholarly debate. In 1942, Paul Benedict first linked Tai languages to a small group of languages spoken on the island of Hainan and in southwestern China, for which he coined the term 'Kadai'. Whereas the Tai-Kadai link is, today, accepted by many – but not all – linguists, Benedict's attempt to relate Tai-Kadai languages to the polysyllabic, nontonal, Austronesian (or Malayo-Polynesian) languages of the South Pacific, under the term 'Austro-Tai,' has proved more controversial.

History

Researchers on comparative Tai dialects estimate that the parent language, Proto-Tai, dates back approximately 2000 years. Speakers of this language were once thought to have originated in China and migrated southward, but today the border area between Vietnam and China's Guangxi province is regarded as a more likely origin. From the 8th century A.D., Tai speakers began to migrate westward and southwestward, gradually driving a wedge between the Mon-Khmer speaking peoples then dwelling in what is now Thailand. Around the 11th century, all Tai languages were affected by the Great Tone Split. Essentially, this had the effect of creating additional tones while reducing the number of initial consonant sounds. The effects of this can be seen in a number of Tai writing systems; in Thai, for example, it accounts for the fact that a single tone mark can represent two distinct tones.

Typological Characteristics

The Tai languages are noninflected tonal languages with a basic monosyllabic lexicon. Among the different branches, a single lexical item will often show differences in the initial consonant, vowel, or tone; thus, the word for 'six' is *hok* in several southwestern Tai languages, but is *sok*, *rok*, *lok*, or *huk* in the central and northern branches, depending on the language. Even closely related languages within the same branch are frequently mutually unintelligible because of differences in phonology and certain basic vocabulary items.

The word order in most Tai languages is subject-verb-object, with adjectives following nouns. In Khamti, however, the order is subject-object-verb, probably due to the influence of neighboring languages from other families. Geographical location has also influenced the source of loan words; Tai languages spoken in Vietnam and China have borrowed from Chinese, and members from the southwestern branch have drawn lexical items from Sanskrit and Pali.

The writing systems have been similarly influenced; some central and northern Tai languages are written in Chinese characters, whereas southwestern Tai languages are written in alphabetic scripts that can ultimately be traced back to a south Indian origin. Many Tai languages, however, have no writing system and, with small numbers of speakers and little cultural prestige attached to them, they are in serious danger of becoming extinct.

Bibliography

- Benedict P K (1975). *Austro-Thai: language and culture, with a glossary of roots*. New Haven, CT: HRAF Press.
- Computational Analyses of Asian and African Languages 6 (October)* (1976) (much of this issue is devoted to debate on Benedict's Austro-Thai hypothesis).
- Diller A V N (in press). *The Tai languages*. London: Routledge.
- Edmondson J & Solnit D (eds.) (1997). *Comparative Kadai: the Tai branch*. Dallas, TX: Summer Institute of Linguistics/University of Texas at Arlington.
- Gething T W, Harris J G & Kullavanijaya P (eds.) (1976). *Tai linguistics in honor of Fang-Kuei Li*. Bangkok, Thailand: Chulalongkorn University Press.
- Harris J G & Chamberlain J R (eds.) (1975). *Studies in Tai linguistics in honor of William J. Gedney*. Bangkok, Thailand: Central Institute of English Language, Office of State Universities.
- Huffman F E (1986). *Bibliography and index of mainland Southeast Asian languages and linguistics*. New Haven, CT: Yale University Press.
- Kalaya Tingsabath M R & Abramson A S (eds.) (2001). *Essays in Tai linguistics*. Bangkok, Thailand: Chulalongkorn University Press.
- Li F-K (1977). *A handbook of comparative Tai*. Honolulu, HI: University Press of Hawaii.
- Smalley W A (1994). *Linguistic diversity and national unity: language ecology in Thailand*. Chicago, IL: University of Chicago Press.
- Strecker D (1990). 'Tai languages.' In Comrie B (ed.) *The major languages of East and South-East Asia*. London: Routledge. 19–28.

Tajik Persian

J R Perry, University of Chicago, Chicago, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Tajik Persian (self-designation (*zabon-i forsi-i tojikī*; also called Tajik, Tajiki, Tojikī, and Tadjhik) is the variety of Persian used in Central Asia (see **Persian, Modern; Persian, Old**). Since the 1920s, Tajik has been fostered as the national literary language of the Tajik Soviet Socialist Republic (since 1991, the Republic of Tajikistan). It is also spoken in parts of Uzbekistan (notably the cities of Bukhara and Samarkand) and is the vernacular of the Bukharan Jews. It is the common written language and contact vernacular in the mountain region of Badakhshan, where people speak a variety of very different Iranian languages (see **Iranian Languages**). The so-called Tajiks of southwestern Xinjiang in China speak Sarikoli and Wakhi, not Persian. Tajik has been written in a modified Cyrillic script since 1940. Speakers number at least 5 million.

History

Persian spread to Central Asia from its home on the Iranian plateau during the 8th century C.E., as the language of Iranian converts attached to the invading Arab Muslim armies. At the autonomous Samanid court of Bukhara (9th–10th centuries), Persian was patronized as the literary language and displaced the indigenous Iranian language, Sogdian (a descendent of which, Yaghnobi, survives in the mountains of western Tajikistan). As a written language, Persian of Central Asia was hardly distinguishable from Classical Persian of Iran, Afghanistan, and India up until the early 20th century. However, invasions and settlement by Turkic peoples (most recently, the Uzbeks) in the Oxus basin and its foothills interrupted the dialect continuum; spoken Persian of Central Asia evolved independently of Persian of Iran, and northern dialects in particular were strongly influenced by Turkish speech. Persian speakers of the region came to be called Tajiks (from a Middle Persian word meaning ‘Arab’), in contradistinction to Turks.

After the Russian revolution, in accordance with Soviet nationalities policy, an ethnic Tajik republic was established and a literary language called ‘Tajik’ was engineered on a vernacular base close to the Uzbekized spoken Persian of Bukhara and Samarkand (these Tajik cultural centers, ironically, were incorporated in the Uzbek Soviet Socialist Republic). During the period ~1948–1988, Tajik lost much of its prestige, vocabulary, and domain of use, to Russian.

With *perestroika* and *glasnost* came a revival and re-Persianization of the national language, which continues (at a slower rate) in post-Soviet Tajikistan; policies include the replacement of Russian vocabulary by Persian (both native coinages and loans from Persian of Iran), and teaching of the Perso–Arabic writing system in schools.

Tajik is fundamentally Persian in grammar and core vocabulary, though generally closer to the spoken Persian of Afghanistan (e.g., Kaboli dialect) than to Standard Persian of Iran. The following descriptions highlight features that differ substantially from Standard Persian, in particular the elements of convergence with Turkic types characteristic of the bulk of Tajik literature in the Soviet period.

Phonology and Orthography

The Tajik sound system is shared almost entirely with that of Uzbek. Its Cyrillic orthography is basically Russian specific, and is illustrated here only when it involves modified or ambiguous characters.

The consonant inventory differs from that of Persian only in two features: [q] <Қ> and [ɣ] <Ғ> are distinct phonemes (they have collapsed in the Persian of Iran), and labiodental [v] tends toward bilabial [β] or [w] in the environment of rounded vowels. The affricates [tʃ] <ҷ> and [dʒ] <ҷ> will be transliterated as *č* and *j*, [j] <й> will be transliterated as *y*, and <х> [h] will be transliterated as *h*.

The six-vowel system has diverged considerably from Standard Persian (see **Figure 1**). Length has been neutralized in most dialects (including literary Tajik) and replaced by a contrast between ‘stable’ [e, ʊ, ɒ] and ‘unstable’ vowels [i, u, a]. In Cyrillic, [ū] is written *ӯ* (transliterated *ū*); *i*, written as *и*, has a variant *ӣ* (transliterated *ī*). These accents do not represent length: *ū* shows a different quality from *u*, and *ī* is used for *i* in word-final position to distinguish a (stressed) morphological syllable from the (unstressed) enclitic of *izofat* (see later, Morphology and Noun Phrase Syntax). The vowel [e] represents early New Persian [eː]; [ʊ], sounding between [u] and [y], represents early New Persian [oː] and is shared with Uzbek, in which it corresponds to Turkic [y] or [ø]. The vowel [ɒ] <о> is a rounded form of Standard

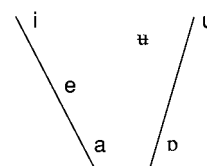


Figure 1 Tajik vowels.

Persian [a]. The three ‘unstable’ vowels correspond to the ‘short’ vowels of Persian, but [i] and [u] additionally represent the corresponding ‘long’ vowels.

Examples of modern correspondences in Tajik and Persian, respectively, are *kitob*, *ketâb* ‘book’; *imrûz*, *emruz* ‘today’; *Bedil*, *Bidel* ‘name of a poet’; *na-budem*, *nabudim* ‘we were not’; and *šuda*, *šode* ‘having become’ (the final [a] is not raised in Tajik). The yotated letters ё, ю, and я represent the syllables *yo*, *yu*, and *ya*; Cyrillic *e* stands for *e* after a consonant, *ye* initially or after a vowel.

Morphology and Noun Phrase Syntax

There is no grammatical gender in Tajik and only a limited distinction between humans and nonhumans in the plural suffixes *-ho* (any noun) and *-on* (humans and higher animals), and in third-person singular personal pronouns, as follows: *vay* (general), *û* (literary), *in* (dialect) ‘he, she’; *on* (literary), *in*, [*h*]*amin* (colloquial), *vay* (dialect) ‘it’; *onho* (general), *in*[*h*]*o* (colloquial), *vay*[*h*]*o* (dialect) ‘they’ (all classes). The deferential pronoun *ešon* (cf. Persian *išân*) ‘he, she’ (lit. ‘they’) has been replaced in Tajik by *in kas* ‘this person.’ Plural pronouns, which may refer deprecatingly or deferentially to a singular (SG) person, can add plural (PL) suffixes as ‘explicit plurals’: *mo*/*mo-yon*, *mo-ho*, *mo-hon* ‘we, I/we’; *šumo*/*šumo-yon*, *šumo-ho* ‘you’ (SG)/‘you’ (PL; see later, discussion of verb endings).

The basic noun phrases (NPs) are the nominal *izofat* (IZ; Persian *ezâfe*), e.g., *qišloq-i Alijon* ‘Alijon’s village’ (village-of Alijon), and adjectival *izofat*, e.g., *qišloq-i kalon* ‘the big village’ (village-big); in both types, the head is linked to a following modifier by the enclitic *-i*. There are no articles; an indefinite NP may be marked by the numeral *yak* ‘one’ and/or the ‘specific’ (SPEC) enclitic *-e*; a definite NP (supplying old information) is distinguished only in the object (OBJ) position, by the enclitic *-ro*. A direct object that is familiar to the speaker, but not to the listener, is marked by both enclitics, as shown in the following examples:

pisar-ro did-am
boy-OBJ see.PAST-1SG
‘I saw the boy.’

yak pisar(-e) did-em
one boy(-SPEC) see.PAST-1PL
‘We saw a boy/some boy or other.’

pisar-e-ro did-em
boy-SPEC-OBJ see.PAST-1PL
‘We saw a (certain) boy.’

Other case relations are expressed through prepositions (including *bar* ‘upon’ and *be* ‘without,’ no longer

active in Persian), postpositions, and circumpositions (inflectional suffix *izofat*):

qayčī kati noxun girift-am
scissors with nail take.PAST-1SG
‘I cut my nails with scissors.’

az ibtido-i paxta-činī in-taraf
from start-IZ cotton-picking this-side
‘Since the start of cotton-picking.’

A superlative as modifier may precede the head noun (as in Persian), or may follow it:

šahr-i kalon-tarin-i tojikiston
city-IZ larg-est-IZ Tajikistan
‘The largest city of/in T.’

Nouns take the singular after a number. A classifier may intervene, most commonly the enclitic *-ta* or *-to* ‘fold, item,’ as in *yak-ta zan* ‘one woman’ and *sad-to kurta* ‘a hundred shirts’ (or *yak-sad kurta* ‘one hundred shirts’).

The simplex tenses of Tajik verbs are the same as in Persian, except for the vowel of the present/imperfect prefix, and of the first-person plural and second-person plural personal endings, as in *me-kun-em* ‘we do’ and *kard-ed* ‘you did.’ The second-person plural form may also add an ‘explicit plural’ supplement (cf. preceding discussion of pronouns) derived from the pronominal enclitic *-ton*, as in *šin-eton*, *rafiq-on* ‘sit down, friends’ (*šined* + *ton*). In compound tenses and moods, Tajik verbal morphology has expanded beyond that of Persian. Three progressive tenses are formed on the past participle of a desemanticized *istodan* ‘to stand’ (in the following examples, verb glosses in bold type indicate an apparent ‘past participle’ (PP) not forming part of a tense, which is used extensively as a nonfinite verb form (gerund) in verbal conjuncts):

bača-ho ovoz xonda istoda-and
child-PL song **sing** stand.PP-be.3PL
‘The children are singing’ (present progressive).

An epistemic mode of the indicative (called ‘nonwitnessed,’ or ‘evidential’) also has three tenses. Thus, the regular perfect may function as an evidential present:

vay sayohat-ba rafta-ast
he journey-on go.PP-be.3SG
‘He went/has gone on a trip (– so I surmise/am told).’

Note here the Persian preposition as a Turkish-style postposition. This mode also includes progressive tenses:

šumo yak asar-i nav navišta istoda-buda-ed
you one work-IZ new **write** stand.PP-be.PP-2PL
‘You’ve been writing a new work (– so I gather/see).’

Here the form expresses a mirative, i.e., the appreciation of a fact not previously known.

The conjectural mood uses an augmented (AUG) form of the past participle in *-agī* to form tenses expressing a probable situation or event (IMPERF, imperfect):

yagon kor-i ganda kardagi-st
some deed-IZ bad do.PP-AUG-be.3SG
'He must have done something bad' (past).

dast-u rū me-šusta-gi-st-ed
hand-and face IMPERF-wash-AUG-be-2PL
'(I imagine) you'll want to freshen up' (present/future).

The future participle (infinitive + adjectival formative *-ī*) is used in a quasifuture tense, and adjectivally, much more than in Persian:

xohar-am ba maktab omad-an-ī bud
sister-my to school come-INF-ADJ be.PAST
'My sister was eager to go to school.'

From an intransitive verb, the sense is active, and with a human subject, usually connotes intention. From a transitive verb, the sense may be passive (NEG, negation):

jo-ho-i no-guft-an-ī
place-PL-IZ NEG-say-INF-ADJ

'Unmentionable places; locations not to be divulged.'

The augmented past participle (also of progressive tenses) is extensively used in ways (and positions, i.e., preceding the head) similar to use in Uzbek participles, to express what, in Persian, would often be a relative clause:

gurexta-istoda-gi-ho
flee-stand.PP-AUG-PL
'Those who are/were fleeing; the fugitives.'

ana kitob-i ovarda-gi-am
here book-IZ bring.PP-AUG-my
'Here is the book that I brought.'

duxtar kurta-i me-dūxta-gi-aš-ro
girl shirt-IZ IMPERF-sew.PP-AUG-her-OBJ
ba modar-aš nišon dod
tomother-her sign give.PAST
'The girl showed the shirt that she was/had been sewing to her mother.'

The Lexicon

Nominal and adjectival compounds are formed with suffixes and prefixes, some of them different from (or more productive than) their Persian counterparts. Thus *-nok* denotes something having the quality of the base noun, as in *foida-nok* 'beneficial, profitable' (*foida* 'use, profit') and *sado-nok* 'vowel' (*sado* 'sound, voice'); *ser-* 'sated, full' indicates an abundance of the base noun, as in *ser-gap* 'garrulous'

(*gap* 'talk'), and *to-* 'up to, until' produces, e.g., *to-inqilob-ī* 'prerevolutionary' (*inqilob* 'revolution,' *-ī* is the relative adjective formative); this use of the preposition *to* (unknown with Persian *tâ*) is probably calqued on similar use of Russian *do* 'up to, until.' Other Russian calques use Tajik *sar* 'head' by analogy with the Russian prefix *glav-*, as in *sar-muhandis* 'chief engineer' (Russian *glav-inžener*). Most derivatives from Russian loans freely use Tajik Persian formatives, as in *bolševik-ī* 'Bolshevik' (adjective).

Transitivizing denominal verbs and causatives (CAUS) (obtained by infixing *-on-*) are more productive than in Persian, as in, *kollektiv-on-idan* 'to collectivize.' They may also be formed from complex and composite verbs:

papiros dar me-gir-on-ad
cigarette in IMPERF-take-CAUS-3SG
'She lights a cigarette.'

(Compare *dar me-gir-ad* 'it catches fire.') In some complex verbs, the preverbs *dar* and *bar* are attached to the verb stem: *me-dar-o-y-ad* 'he comes in' (cf. Persian *dar mi-â-y-ad*).

Characteristic of Tajik are conjunct verbs (serial verbs), of which the progressive tenses are grammaticalized instances. There are some 18 lexically established conjunct auxiliaries (corresponding to models in Uzbek) that, in regularly conjugated tenses, furnish adverbial 'modes of action' for the nonfinite participle (semantically, the main verb):

dars-i nav-ro navišta girift-em
lesson-IZ new-OBJ write take.PAST-1PL
'We copied down the new lesson' ('take': self-benefactive).

nom-i xud-ro navišta me-dih-am
name-IZ own-OBJ write IMPERF-give.PRES-1SG
'I'll jot down my name (for you)' ('give': other-benefactive).

berun-ho-ya toza kardā rūfta partō!
outside-PL-OBJ clean make sweep throw.IMP
'Sweep all the outside nice and clean!'

The preceding example demonstrates a double conjunct construction: the auxiliary *partoftan* 'to throw (away), toss' adds the sense of thoroughness or completion (*-ya* is a dialect variant of *-ro*, and *toza kardan* 'to clean' is a typical Persian-type composite verb).

Syntax

Verbal conjuncts, mostly of Uzbek inspiration, compete in other ways with the Persian syntax of subordinate clauses introduced by conjunctions; e.g., the favored construction for the modal verb *tavonistan* 'to be able' is as follows:

man rafta (na-)me-tavon-am
 I go (NEG-)IMPERF-can.PRES-1SG
 'I can(not) go.'

Also embedded in the literary language is the nominalization of sentential complements through infinitives, as in the following example:

mo {kujo raft-an-i xud-ro}
 we {where go-INF-IZ own-OBJ}
 me-don-em
 IMPERF-know.PRES-1PL
 'We know where we are going' (... our going-where).

Uzbekisms in colloquial and northern dialect usage include the question (Q) enclitic *-mi* and possessive NPs, with (dative) *-ro* replacing the *izofat* construction (OBL, oblique):

muallim-a [-ro] pisar-aš raft-mi?
 teacher-OBL boy-his go.PAST-Q
 'Has the teacher's son left?'

These features were not admitted into literary Tajik, and even some of the accepted Uzbekisms are fading from post-Soviet Tajik writing.

Bibliography

- Baizoyev A & Hayward J (2004). *A beginner's guide to Tajiki*. London: Routledge Curzon.
 Bashiri I (1994). 'Russian loanwords in Persian and Tajiki languages.' In Marashi M (ed.) *Persian studies in North*

- America: studies in honor of Mohammad Ali Jazayery*. Bethesda, MD: Iranbooks. 109–139.
 Lazard G (1956). 'Caractères distinctifs de la langue Tadjik.' *Bulletin de la Société de Linguistique de Paris* 52/1, 117–186.
 Perry J R (1996). 'From Persian to Tajik to Persian: culture, politics and law reshape a Central Asian language.' In Aronson H I (ed.) *NSL.8. Linguistic studies in the non-Slavic languages of the Commonwealth of Independent States and the Baltic republics*. Chicago: The University of Chicago, Chicago Linguistics Society. 279–305.
 Perry J R (1997). 'Script and scripture: the three alphabets of Tajik Persian, 1927–1997.' *Journal of Central Asian Studies* 2/1, 2–18.
 Perry J R (2000). 'Epistemic verb forms in Persian of Iran, Afghanistan and Tajikistan.' In Johanson L & Utas B (eds.) *Evidentials: Turkic, Iranian and neighboring languages*. Berlin, New York: Mouton de Gruyter. 229–257.
 Perry J R (2005). *A Tajik Persian reference grammar*. Leiden: Brill.
 Rastorgueva V S (1963). *A short sketch of Tajik grammar*. Paper H H (trans./ed.). Bloomington, IA: Indiana University and The Hague: Mouton.
 Rzehak L (2001). *Vom Persischen zum Tadschikischen. Sprachliches Handeln und Sprachplanung in Transoxanien zwischen tradition, moderne und Sowjetmacht (1900–1956)*. Wiesbaden: Reichert.
 Soper J D (1996). 'Loan syntax in Turkic and Iranian: the verb systems of Tajik, Uzbek and Qashqay.' In Bodrogligeti A J E (ed.) *Eurasian language archives 2*. Bloomington, IN: Eurolingua.

Tamambo

D Jauncey, Australian National University,
 Canberra, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Tamambo [*tama^mbo*] (Malo) is the predominant dialect of the language of the island of Malo (previously known as St. Bartholomew) in northern Vanuatu, in the southwest Pacific (Figure 1). It is spoken by at least 3000 people including those living on Malo, and those who have settled on the nearby 'big' island of Espiritu Santo and in Port Vila. It is learned as a first language by most children on the island, although Bislama (Vanuatu pidgin) is strengthening in almost all social contexts. Tamambo was originally the dialect of the western side of Malo; the dialect of the east

[*tamapo*] is now used by no more than a handful of older speakers, although some words from that dialect are heard in several old dance songs. There is no written literature in the language, except for some copies of Presbyterian mission publications dating from the 1890s. Nevertheless, a strong oral tradition of storytelling has been maintained, and activities reflecting *Kastom* (traditional custom) such as dances, and 'fighting sticks' contests [*maⁿ'ja*] are enjoying renewed interest and participation.

Grammatical Overview

The language is Oceanic (Austronesian); it belongs to the Northern Vanuatu linkage, and appears similar to languages of nearby Tangoa, Araki, and south Santo. Tamambo can be regarded as conservative in that it shares many of the same structural characteristics

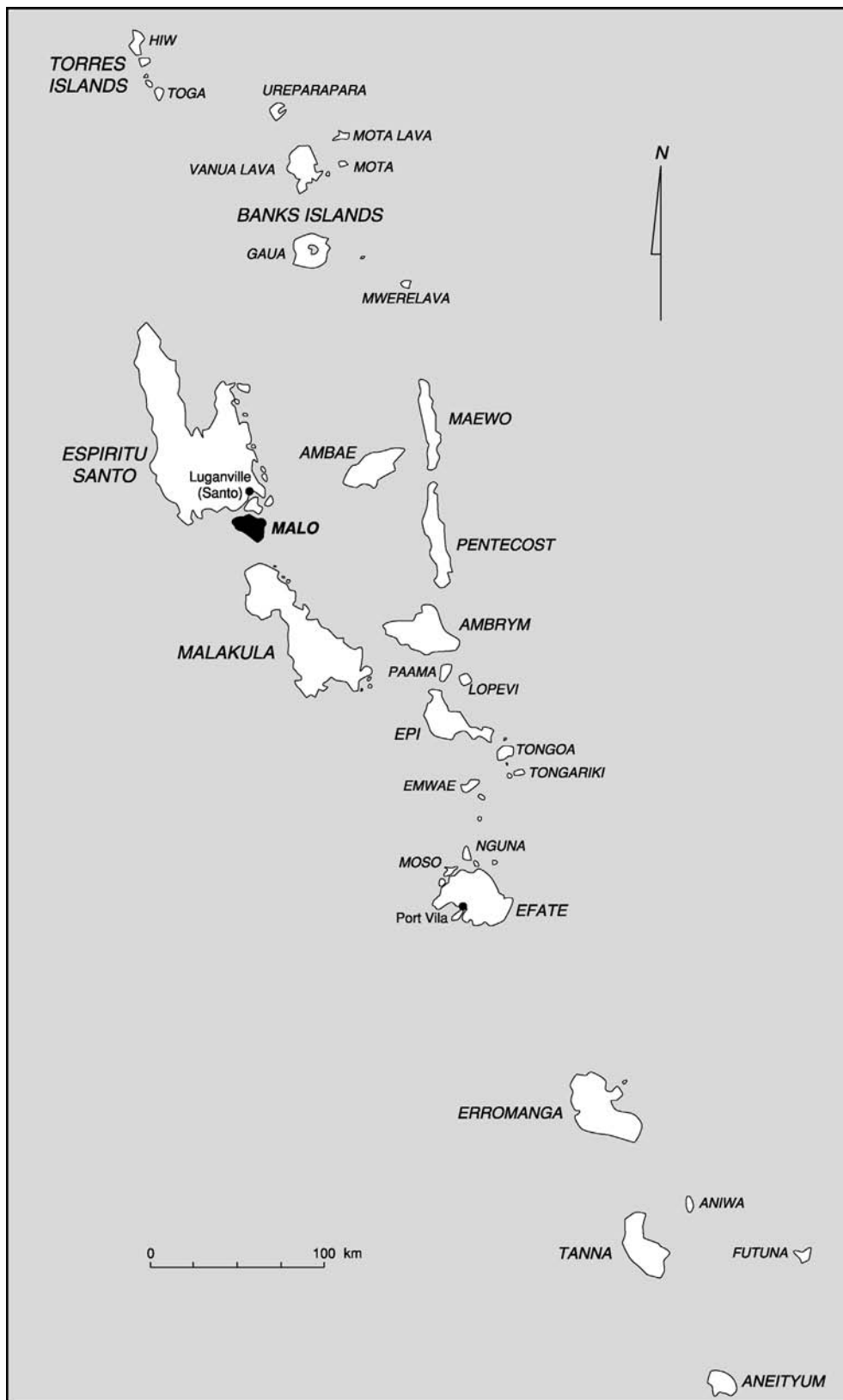


Figure 1 Malo Island within Vanuatu.

widely distributed among Oceanic languages, and many of which are posited for Proto-Oceanic (POc).

Tamabo is a nominative-accusative language, and the unmarked word order of the clause is Agent-Verb-Object or Subject-Verb. Sentence types other than the declarative are based on the unmarked declarative form. Basic clauses are most commonly verbal clauses that indicate a non-future/future contrast. There are also verbless clauses where the predicate is a noun phrase, a numeral, or a prepositional phrase. Basic noun phrase structure is similar to that outlined for POc (Lynch *et al.*, 2002: 75) with the noun as head, preceded by an article (retained only in some syntactic environments in Tamabo), and an optional premodifier such as a quantifier, and followed by an optional modifier or demonstrative. It is an agglutinating language with considerable derivational morphology and valency-changing affixes.

Lexically, many words in the language are reflexes of words posited for POc. Other characteristics common to many Oceanic languages are reflected in Tamabo: they include a subject proclitic on the verb root, marking of inclusive and exclusive distinctions in pronouns, spatial concepts 'seaward' vs. 'inland,' 'up direction' vs. 'down direction' (depending on location on island) indicated by particular verbs and/or location nouns, possessive constructions with noun phrases reflecting the semantics of alienable and inalienable possession, and 'tail-head' linkage of clauses in procedural narrative.

Phonology

Tamabo reflects many of the consonants of the reconstructed POc paradigm (Lynch *et al.*, 2002: 63) with little or no phonetic change. Voiced stops are prenasalized.

| Bilabial | Dental-alveolar | Pre-palatal | Velar |
|---------------------|-----------------|-------------|-------|
| b b ^w | t d | ʃ | k |
| m m ^w | n | | ŋ |
| β β ^w | s | | x |
| | r | | |
| | l | | |

Like POc, there is a five-vowel system, sequences of unlike vowels are permitted, and syllable structure is primarily (C)V.

Orthography

The four prenasalized stops are written as b, bw, d, j. Fricative /β/ is written as v, the additionally labialized fricative as w; /x/ is represented by h, /m^w/ by mw and /ŋ/ as ng.

Examples of Particular Grammatical Characteristics

Productive Derivations from Affixation, Reduplication, and Compounding

Affixes to nouns

| | | |
|--------------------------------|--------------------|-------------------------------------|
| -ha <i>noun-like quality</i> | dodo → dodo-ha | 'night' → 'be cloudy/dark' |
| -a <i>nominalization</i> | luhu → luhu-a | 'hide' → 'refuge' |
| vo- <i>female</i> | natuku → vo-natuku | 'my child/son' → 'my daughter' |
| ta- <i>person belonging to</i> | ta-Alotu | 'Santo person' |
| vu- <i>tree</i> | vu-talaua | 'sago palm tree' |
| lo- <i>plural (trees only)</i> | lo-vu-talaua | 'sago palm trees' |
| ra- <i>female plural/leaf</i> | ra-vavine | 'women'; ra-talaua 'sago palm leaf' |

Reduplication of nouns and verbs

| | |
|---------------------|---|
| hinau → hina-hinau | 'thing' → things |
| mata → mata-mata | 'eye' → 'signs' |
| bange → bange-bange | 'stomach' → 'pregnant' |
| mana → mana-mana | 'laugh' → 'friendly' |
| sahe → sahe-sahe | 'go up' → 'keep going up' |
| tau → tau-tau | 'put s.t in place' → 'put many things in place' |

Compounding (noun + noun; noun + verb; verb + verb)

| | |
|------------|------------------------------------|
| mara-rohai | 'man-leaf/leaves' → 'medicine man' |
| mata-suri | 'eye-follow' → 'be jealous' |
| bosi-mate | 'turn-die' → 'extinguish (lamp)' |

Valency Changing Affixes

| | |
|--------------------|--|
| -hi, -si | 'applicatives'; ma- 'agentless passive'; |
| va-/vaha- | 'causative', and vari- 'anti-passive' |
| sora → sora-hi | 'talk' → 'talk about s.t.' |
| lua → lua-si | 'vomit' → 'vomit on s.t.' |
| duru → ma-daru | 'split s.t.' → 'be split' |
| mauru → vaha-mauru | 'be alive' → 'save life' |
| hati → vari-hati | 'bite s.t.' → 'inclined to bite' |

Serial Verb Constructions for a Variety of Functions

Action in specified direction

| | | |
|-----------------|-------------|----------------|
| vavine | le-hilo | le-sahe |
| woman | ASP-look | ASP-go.up |
| ta-vonavu | mo-dono | mo-jivo |
| belong-Malakula | 3.sing-sink | 3.sing-go.down |
| ana | tarusa | |
| PREP | sea | |

'while the woman was looking up, the Malakula man drowned in the sea'

Comparative

| | | |
|-------------------------------------|-------|---------------|
| heletu | niani | mo-suiha |
| pig | this | 3.sing-strong |
| mo-liu-ra | | |
| 3.sing-win.over-O.3PL | | |
| 'this pig is the strongest of them' | | |

Continuative aspect

ku-vano ku-le ovi, ku-ovi
 1.sing-go 1.sing-ASP stay 1.sing-stay
 mo-vano mo-vano ...
 3.sing-go 3.sing-go
 'I went and I was waiting, I kept on and on
 waiting ...'

Completive aspect

voi mo-mule mo-iso
 mum 3.sing-head.home 3.sing-finish
 'mum has already gone home'

Non-result

ka-te soari-a, ka-sai-a
 1PL-NEG see-OBJ.3.sing 1PL-search-OBJ.3 sing
 mo-tete
 3.sing-negative
 'we didn't see it, we searched for it to no avail'

Possessive Constructions*Classifiers for inalienable possession*

no- personal property
 no-da vanua 'our (INCL) house'
 ma- drinkable
 ma-m reu 'your (sing.)water (to drink)'
 ha- edible
 ha-mam vetai 'our (EXCL) bananas'
 bula- living things (animals, crops + things regarded as 'living')
 bula-ra toa 'their chickens'; bula-ku radio 'my radio'

Overlap between constructions or classifiers

no-ku nunu 'my photo' (that I own)
 nunu-ku 'my photo' (of me)
 bula-na dam 'his yam/s' (growing)
 ha-na dam 'his yam/s' (to eat)

Hierarchy of Individuation: Kin Terms/Proper Names → Animate → Inanimate*Differentiation of kin/proper names vs. common nouns*— *comitatives* mai/mana

Voi mai Alis vavine atea mana mwera atea
 'Mum and Alice' 'a girl and a boy'

— *possessive linkers* ni/i indicating 'status' of possessor
 naho-ni vuti-ni Abae tamanatu-i
 mama vavine ridi
 face-POSS hill/s-POSS Ambae. husband-POSS
 dad woman DEF
 'dad's face' 'the hills of Ambae' 'the woman's
 husband'

— *prepositions* hini/ hina

hini Air Vanuatu hina siba
 'with Air Vanuatu' 'with a knife'

Differentiation of animate vs. inanimate— *quantitative verbs*

tamalohi na-were heletu na-were sala mo-were
 person 3PL- pig 3PL- road 3sing-
 be.many be.many be.many
 'many people' 'many pigs' 'many roads'

— *prepositions* telei/ana

telei-au telei bula-ku vuria ana tano
 'to me' 'to my dog' 'to the garden'

Bibliography

- Crowley T (1987). 'Serial verbs in Paamese.' *Studies in Language* 11(1), 35–84.
 Durie M (1997). 'Grammatical structure in verb serialization: some preliminary proposals.' In Alsina A, Bresnan J & Sells P (eds.) *Complex predicates*. Stanford, CA: CSLI Publications. 289–354.
 Lichtenberk F (1985). 'Possessive constructions in Oceanic languages.' In Pawley A & Carrington L (eds.) *Austronesian linguistics at the 15th Pacific Science Congress*. Canberra: Pacific Linguistics. 93–140.
 Lynch J & Crowley T (2001). *Languages of Vanuatu: a new survey and bibliography*. Canberra: Pacific Linguistics.
 Lynch J, Ross M & Crowley T (2002). *The Oceanic languages*. Richmond, Surrey: Curzon.

Tamil

H F Schiffman, University of Pennsylvania,
 Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Tamil is the Dravidian language with the most ancient literary tradition in India, dating from the early centuries A.D. or before. The earliest (3rd–1st century B.C.) inscriptions of Tamil are found in caves used

by Buddhist and Jain monks, in a form known as Tamil *Brahmi* script. The earliest text in Tamil is a grammar, the *Tolkaappiyam*, which describes *centamiṟ* (Old Tamil) with both literary and colloquial (*koDuntamiṟ*) dialects, spoken in what is now Tamilnadu and Kerala, in South India. An early and original poetic literature, known as *Sangam* Tamil, has survived in the form of various anthologies; these early texts show few borrowings from Sanskrit, and

minimal Brahmanic or 'Hindu' influences. After Old Tamil, a Middle Tamil literature can be distinguished, marked by diverse influences, including increasing Aryanization, Buddhism, and Jainism. Two epics, the *Cilappatikaram* (The lay of the ankle bracelet) and *Maṇimekalai* (The girdle of jewels), a Buddhist work, date from the 4th–6th centuries A.D., and the *Tirukkural*, known to every Tamil and considered by many to be the apex of their literary genius. In the 6th–9th centuries, *bhakti* devotional poetry (the hymns of the Alvars and Nayanars), devotional literature honoring Vaishnava and Shaiva saints, developed, then spread as a phenomenon across India.

From this period until the arrival of Western colonizers and missionaries, Tamil literature reflects pan-Indian norms devoted to philosophical and religious writings, with little originality (and heavy Sanskritization) except for the poetry of Kampan. After the consolidation of colonialism, Tamil literature shows more influences of Western, especially English, ideas. But the development of English education in India also stimulated resistance to these norms and a renaissance and revival of Tamil, focusing on purifying the language of Indo-Aryan and other loan words.

The Tamil language has had its current standard written form since the thirteenth century, when codified again in the grammar *nannuul*, composed (according to some accounts) by the Jaina monk Pavanandi. But due to increasing diglossia (Britto, 1986), spoken Tamil dialects have now diverged so radically from earlier norms, including the written standard (LT, or Literary Tamil; Arden, 1942) that no spoken dialect (regional or social) can function as the koiné or lingua franca. Since LT is never used for authentic informal oral communication between live speakers, there has always been a need for some sort of spoken 'standard' for inter-dialect communication, and what has evolved has been hastened by the development of modern communication, especially the 'social' film, which is the chief disseminator of this 'standard spoken Tamil' (SST). This form (Schiffman, 1999), based on the everyday speech of educated non-Brahman Tamils, is understood wherever Tamil is spoken, including Sri Lanka, Malaysia, and Singapore.

The sound system of Tamil consists of a ten-vowel system with long and short *i* and *iː*, *e* and *eː*, *a* and *aː*, *o* and *oː*, and *u* and *uː*. The diphthongs *ai* and *au* are found in LT but are not usual in ST; a few loan words contain *au*, but often these can be represented by *avu* as in *pavuṇḍu* 'pound.' The vowel *u* has an unrounded variant [u] that occurs after the first syllable, and there are also nasalized variants [ã], [õ], as well as nasalized versions of [ẽ] and [ũ], all found in

final position only (as the result of deletion of final nasals in SST, but not in LT).

In LT, as in Proto-Dravidian, there was a series of six stop consonants: velar *k*, palatal *c*, retroflex *ʈ*, alveolar *t̪*, dental *t*, and labial *p*. The apical stops *ʈ* and *t̪* could not occur in initial position. In non-initial position, all stops were voiced after nasals (i.e., they were phonetically *g*, *j*, *ɖ*, *ɖ̪*, *d*, and *b*), and intervocalically, unless geminated, they were lax (i.e., phonetically *h*, *s*, flapped *ɾ*, flapped *r*, *ð*, and *ʋ*). Since these variants are in complementary distribution, no contrast between voiced and voiceless consonants (and the fricative variants) existed. In modern SST, because of borrowings, voiced consonants occur in other environments than these, so the phonological system now has voiced stops, though the orthography lacks provisions for this.

Today because of the loss of the alveolar contrast, modern SST only has contrasts between five points of articulation in consonantal stops, with voiced variants in many loan words (but also in onomatopoeic expressions, of which there are many). Nasal consonants (despite orthographic symbols for all six positions) are only *m*, *n*, and retroflex *ɳ*. In the area of laterals and rhotics, there is confusion. Proto-Dravidian surely had contrasts between *l* and retroflex *ɭ*, and *r* and *ɻ* a 'retroflex frictionless continuant' symbolized variously in transcriptions, but for which we prefer *ɻ*, but because of the loss of the intervocalic alveolar stop contrast (*t̪*), which is flapped [ɾ] in modern speech, orthographic symbols for three *r*'s exist. Furthermore, the retroflex continuant *ɻ*, which happens to be the final segment in the name 'Tamil' (*tami.ɻ*) is often not maintained in speech in many dialects, merging instead with *l*, *g*, and even *y*. But sociolinguistic pressure to maintain this sound, seen as quintessentially *Tamil*, results in much variation in its maintenance. As for glides, both *y* and *ʋ* (which varies sometimes to [w]) are found.

Grammatically, Tamil can be characterized as 'agglutinative,' with long chains of easily-identifiable morphemes concatenated as suffixes. Noun morphology is fairly simple (there is no grammatical gender), and noun phrases require no agreement with adjectives and nouns. A seven-case system with a nonfinite set of postpositions recruited from lexical items both nominal and verbal, completes the picture. Example:

anta periya viiṭṭ-ukk-pakkattu-le- rundu
That large house-DAT-near LOC + ABL
'From the vicinity of that large house'

The verbal system is morphologically more complex, with various inflectional and derivational morphemes concatenated as suffixes. Example:

avarai eppaiyoo anuppu- vittu-viṭa- veenṭum
 he-ACC somehow send CAUS-COMPL-MODAL
 'Somehow or other, (we) have to get rid of him'

Syntactically, word order is SOV and left-branching. Grammaticalization processes have resulted in the incorporation of certain lexical verbs into the morphology of the verb as 'aspectual' markers, a phenomenon typical in many South Asian languages.

Tanoan

L J Watkins, Colorado College, Colorado Springs, CO, USA

© 1994 Elsevier Ltd. All rights reserved.

The majority of Tanoan-speaking peoples have inhabited pueblos in the American Southwest for at least two thousand years. Only the Kiowas are plains dwellers, having occupied the southern plains for about two hundred years.

Subgroups, Locations, and Speakers

The Tanoan languages fall into four subgroups which show varying degrees of internal diversity.

Tiwa consists of two languages, separated geographically by the Tewa-speaking pueblos. Northern Tiwa comprises two very divergent dialects spoken at the northern New Mexico pueblos of Taos, with perhaps 1000 adult speakers, and Picuris. Southern Tiwa, whose varieties differ only slightly, is spoken at the pueblos of Isleta and Sandia, located in the vicinity of Albuquerque. Numbers of fluent adult speakers range from about 2000 at Isleta to fewer than a dozen elderly individuals at Sandia.

The major dialect division in Tewa reflects the emigration of Tewas usually identified as Tanos from the Rio Grande area at the time of the seventeenth-century Pueblo Revolt. Rio Grande Tewa is spoken by roughly 1000 adults at five pueblos clustered just north of Santa Fe, New Mexico: San Juan, Santa Clara, San Ildefonso, Tesuque, and Nambe. These mutually intelligible dialects exhibit only minor phonological and lexical differences. Arizona Tewa (also called Hopi-Tewa) is spoken fluently by approximately 300 speakers (including some children) who live in a multilingual community located at Hopi First Mesa in north-eastern Arizona.

Towa is the language of Jemez Pueblo, located in the Jemez mountains of New Mexico to the west of

Bibliography

- Arden A H (1942). *A progressive grammar of Tamil*. Madras: Christian Literature Society.
 Britto F (1986). *Diglossia: a study of the theory with application to Tamil*. Washington, DC: Georgetown University Press.
 Schiffman Harold (1999). *A reference grammar of spoken Tamil*. Cambridge: Cambridge University Press.

the Rio Grande. It continues to be the first language of Jemez children in a population of approximately 2000.

Kiowa, the only non-pueblo language of the family, is spoken today by perhaps 300 older adults in south-western Oklahoma. Prior to 1700, when ethnohistorical research puts the Kiowas in western Montana, nothing is known of their earlier location or migration.

History and External Relationships

Internal relationships within Tanoan are complex and poorly understood. Although Tiwa and Tewa have been considered to be more closely related than either is to Towa or Kiowa, and Kiowa to be the most divergent, the closer resemblances among the pueblo languages may well be attributable to centuries of contact. Hale and Harris's (1979) proposal that Tanoan consists of four roughly coordinate branches appears to be supported by current comparative work: e.g., phonological innovations show less definitive subgrouping than previously described.

A more distant relationship with Uto-Aztecan, long thought plausible and incorporated in Sapir's Aztec-Tanoan group, remains an open question that has received little recent attention.

Phonological and Grammatical Features

The Tanoan languages have fairly complex phonological inventories. They share a four-way stop contrast of voiceless unaspirated, voiceless aspirated (fricatives in Towa and for some positions in Tiwa and Tewa), glottalized, and voiced. The languages have six vowel qualities, with contrastive nasalization, and for some languages contrastive vowel length. In all four subgroups there is contrastive tone (high, falling, and low). Grammatically, the languages show triple agreement, that is, fused (or portmanteau) verbal prefixes which encode three arguments for person,

number, and case. Verbal morphology includes extensive stem alternation as well as suffixation, ablaut in stem-initial consonants, and incorporation of nominal, verbal, and adverbial roots. Nouns are classified according to animacy and number; plurals of animate nouns and singulars of some inanimate nouns are morphologically alike. Basic word order is verb-final, but nouns may follow the verb depending on discourse context. Tiwa and Towa are noted for unusual passives constrained by a topicality hierarchy.

Future Scholarship

Much of the research on Tanoan languages remains unpublished in dissertation or manuscript form. Hale (1967) provides a phonological survey with discussion of morphophonemic alternations. Grammatical sketches for Northern and Southern Tiwa

are in preparation. San Juan (Tewa) Pueblo has made available a dictionary and collection of stories. Towa, about which the least material is available, is now the topic of two dissertations. For Kiowa, a grammar (Watkins 1984) will soon be supplemented by a dictionary and collection of texts.

Bibliography

- Hale K L (1967). Toward a reconstruction of Kiowa–Tanoan phonology. *IJAL* 33, 112–120.
 Hale K & Harris D (1979). Historical linguistics and archeology. In Ortiz A (ed.) *Handbook of North American Indians, The Southwest*. Washington, DC: Smithsonian Institution.
 Watkins L J (1984). *A Grammar of Kiowa*. Lincoln, NE: University of Nebraska Press.

Tariana

A Y Aikhenvald, La Trobe University, Bundoora, VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

The Tariana language belongs to the Arawak language family (see **Arawak Languages**). It is spoken by about 100 people in the multilingual linguistic area of the Vaupés River Basin (northwest Amazonia, Brazil). This area is known (Aikhenvald, 2002b; Sørensen, 1967) for its multilingual exogamy: one can only marry someone who speaks a different language and belongs to a different tribe. People usually say: ‘My brothers are those who share a language with me’ and ‘We don’t marry our sisters.’ The other languages in this area belong to the Tucanoan family, and they are still spoken by a fair number of people. The basic rule of language choice throughout the Vaupés area is that one should speak the interlocutor’s own language. Descent is strictly patrilineal, and consequently, one identifies with one’s father’s language group. There is a strong cultural inhibition against ‘language-mixing,’ viewed in terms of lexical loans. In its grammatical and semantic structure, Tariana combines a number of features inherited from proto-Arawak, with the areal influences from Tucanoan in the form of grammatical calques and diffused patterns.

Tariana was once a dialect continuum spoken in various settlements along the Vaupés river and its tributaries. The Tariana clans used to form a strict

hierarchy (according to their order of appearance as stated in the creation myth: see Aikhenvald, 1999). Lower-ranking groups in this hierarchy (referred to as ‘younger siblings’ by their higher-ranking tribes people) would perform various ritual duties for their ‘elder siblings.’ Each group spoke a different variety of the language. The difference between these varieties is comparable to that between Romance languages.

As the Catholic missions – and with them white influence – expanded, the groups near the top of the hierarchy abandoned the Tariana language in favor of the numerically dominant Tucano language. This process started in the early 1900s. The Tariana language is spoken nowadays just by people from two subtribes of the lowest-ranking group Wamiarikune, in two villages, Santa Rosa and Periquitos. The varieties are mutually intelligible. Most children are not learning Tariana any more. Innovative speakers of Tariana have more Tucanoan-like features in their language than traditional speakers. A literacy program in Tariana is presently under negotiation.

Tariana is a polysynthetic language, agglutinating with some fusion. It has mostly suffixes, with just a few prefixes. Constituent order depends on pragmatics. Younger speakers tend to put the verb last in the sentence, just like speakers of Tucano. There are mainly postpositions, with just one preposition (borrowed from Portuguese).

Tariana has 27 consonants (including a series of aspirated stops and pre-aspirated nasals and glide)

and 15 vowels (*a*, *i*, *e*, *u*, each with a long and a nasal counterpart), *o* (with a nasal counterpart), and high central *i*. Accent is distinctive and of pitch type, as a result of Tucanoan influence.

Underived adjectives form a closed class of about 30 members, while classes of nouns and verbs are open. Verbs divide into transitive and intransitive active, which take prefixes cross-referencing their subject (A/Sa). As is typical for an Arawak language, the same set of prefixes marks possessor on inalienably possessed nouns and the argument of postpositions. Intransitive stative verbs do not take any cross-referencing markers. Unlike any other Arawak language, grammatical relations in Tariana are also marked with cases: topical non-subject case *-nuku*, focused subject case *-nel/-nhe*, instrumental case *-ine*, and locative case *-se*. This case system for marking core syntactic functions was developed under the Tucanoan influence. The case markers result from the reanalysis of locative suffixes of Arawak origin. A member of any word class can occupy the intransitive predicate slot.

The locative and the instrumental cases can combine with the non-subject topical case if the constituent is topical (thus yielding a peculiar instance of ‘double case’).

Tariana has a complex system of more than 40 classifiers that are used as agreement markers on adjectives, as derivational affixes on nouns, and also as numeral and as verbal classifiers; a slightly different system of classifiers is used with demonstratives. A two-way gender opposition (feminine vs. the rest) is used in personal pronouns (third singular and all plural forms, thus contravening established universals) and in verbal cross-referencing. Classifiers are an open class, since any noun with an inanimate referent can be used as a ‘repeater’ (or ‘self-classifier’). Repeaters can be used to mark the agreement with a topical noun while grammaticized classifiers are used for unmarked agreement.

There is an obligatory distinction between singular and plural for nouns with animate referent. Nouns with inanimate referent often refer to substances, and classifier suffixes are attached to them to specify singular reference. For instance, *episi* means ‘iron as a substance’, while *episi-da* (iron-CLASSIFIER: ROUND) means ‘axe’ and *episi-kha* (iron-CLASSIFIER: CURVED) means ‘wire’. Number agreement is optional for inanimate nouns.

The Tariana verb has a plethora of moods and aspects. It has an elaborate system of marking information source, known as evidentiality. Tariana distinguishes visual evidentials (something seen), non-visual evidentials (something heard, or smelled, or felt by touch), inferred evidential (something inferred based on visible results: as one infers that it

has rained on the basis of puddles); assumed evidentials (based on general knowledge), and reported evidential. Three tenses (present, recent past, and remote past) are combined with evidentials. Traditional stories are typically cast in remote past reported evidential, and autobiographical narratives in visual evidential. Non-visual evidential is used to relate the actions of evil spirits that are not ‘seen’, and dreams of ordinary people, while prophetic dreams by omniscient shamans are cast in visual evidentials. A reduced set of evidentials is used in questions, while imperatives have just one, reported, evidential (meaning ‘do something on someone else’s order’). This unusually complex evidentiality system has been largely calqued from Tucanoan languages.

A complicated system of serial verb constructions expresses aspectual, directional, and sequential meanings, and also reciprocal and associative meanings. There are three types of causatives. Morphological causatives are formed on intransitive verbs. The same morpheme on a transitive verb indicates an advancement of a peripheral argument of the transitive verb to the core, and/or complete involvement and topicality of the O argument. Periphrastic causatives (indirect causation) and serial causative constructions (direct causation) are used to form causatives of transitive verbs.

When several clauses are combined to form one sentence, all but the main clause are marked differently depending on whether their subject is the same as, or different from, that of the main clause. This feature (known as switch-reference) is shared with the Tucanoan languages.

A detailed reference grammar is in Aikhenvald (2003). Aikhenvald (2002a) is a comprehensive dictionary, while Aikhenvald (1999) contains a text collection and an outline of the Tariana ethnography with an account of the kinship system (which is of Dravidian type).

Bibliography

- Aikhenvald A Y (1999). *Tariana texts and cultural context*. Munich: Lincom Europa.
- Aikhenvald A Y (2002a). *Dicionário Tariana-Português e Português-Tariana*. Belém: Museu Goeldi.
- Aikhenvald A Y (2002b). *Language contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald A Y (2003). *A grammar of Tariana, from north-west Amazonia*. Cambridge: Cambridge University Press.
- Sorensen A P Jr. (1967). ‘Multilingualism in the Northwest Amazon.’ *American Anthropologist* 69, 670–684 (reprinted in 1972 in Pride, J B & Holmes, J (eds.) *Sociolinguistics*. Harmondsworth: Penguin Modern Linguistics Readings. 78–93).

Tatar

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Tatar (*tatar tēlē*, *tatarča*) is the designation for Kazan Tatar and related dialects belonging to the northern subbranch of the Northwestern or Kipchak branch of the Turkic language family. It is distributed over a huge area, from Ryazan in the west to West Siberia in the east, from the Kirov region in the north to Astrakhan in the south. Most Tatar speakers live between the Volga–Kama triangle and the western slopes of southern Ural. The Republic of Tatarstan (*Tatarstan Respublikası*) with its capital Kazan is situated in the central part of the Russian Federation, at the confluence of the Volga and the Kama. It borders Bashkortostan in the east, Mari El and Udmurtia in the north, and Chuvashia in the west. This multinational republic has a total population of over 3.8 million, mainly consisting of Volga Tatars (over 50%) and Russians (over 40%). There are also speakers of Chuvash, Mordva, Udmurt, Mari, Bashkir, etc.

Speakers of Mishar Tatar live mainly west, south-west, and south of the republic. Kasimov Tatar was formerly spoken farther west, in the Ryazan region, on the territory of the old Kasimov Khanate. Tatar-speaking groups, often descendants of Noghays, still live along the Volga river south of the republic, down to the Astrakhan region. There are also scattered Tatar-speaking groups in other central parts of the Russian Federation. About 1 million Tatars live in Bashkortostan. The Tepter Tatars live along the Ural River.

East of the Ural Mountains, in an area that was once the home of sizeable Turkic-speaking groups, West Siberian Tatar varieties of different origins are still spoken by small groups, about 150 000 persons altogether: the dialects of Irtysh, Tümen, Tura, Tobol, Tara, Ishim, etc. The Baraba Tatars, about 8000 persons, live in the Baraba steppe, between Novosibirsk and Omsk. Tatar is also spoken in parts of Kazakhstan, Uzbekistan, China, etc. The total number of Tatar speakers is about 8 million.

The designation ‘Tatar’ is ambiguous. Until the end of the 19th century, it was used for all languages of Turkic Muslim groups in Russia. It is still used for Crimean Tatar (Judco Crimean Tatar), which is not identical with Volga Tatar, but a language in its own right. The so-called Tatar minority in China, mostly in Xinjiang (about 5000), consists of descendants of Volga and Crimean Tatars. Groups in Poland, Belarus,

and Lithuania, referred to as ‘Lithuanian Tatars,’ are linguistically assimilated descendants of Noghays and Crimean Tatars. The Tatars of West Siberia partly consist of emigrants from the Volga region. The Baraba Tatars go back to deported Kipchak tribes. The Tatars of Astrakhan and Siberia have strong Noghay elements.

Tatar has long been one of the most firmly established Turkic languages. It has consolidated its position further in the post-Soviet era. The official languages of Tatarstan are Tatar and Russian. Of the Tatars of the Russian Federation, 86% regard Tatar as their mother-tongue.

Origin and History

The designation Tatar is first mentioned in Chinese sources and in Turkic inscriptions of the 8th century. Later on it appears as a Mongol tribal name. Kipchak Turkic groups who arrived in the Volga region with the Mongols adopted it for themselves. It was used for the Turkic and Mongol population in the Golden Horde, also for Turkic groups that arrived later, and finally also for older Turkic groups of the Volga–Kama area.

Tatar is a result of complex linguistic contact processes, the main elements being Kipchak Turkic, Volga Bulgar, Volga Finnic, and Mongolic. Turkic groups were probably present on the middle course of Volga River from the 5th century on, absorbing local Finno–Ugric tribes of the region. The Volga Bulgar element was of decisive importance. The powerful Volga Bulgar state was created at the end the 9th century and adopted Islam in 922. The Volga Bulgars assimilated native groups of the region. Both Tatars and Chuvash regard themselves as descendants of the Volga Bulgars. The state was destroyed by the Mongols in 1237 and the Khanate of the Golden Horde was established. Its most important element was Kipchak Turkic, which became the dominant assimilating factor. Volga Bulgars, Finno–Ugric groups, and Mongols shifted to Kipchak. The speakers of the predecessor of Chuvash, however, were not assimilated but preserved their language. After the disintegration of the Golden Horde, the Khanates of Kazan, Crimea, Kasimov, Astrakhan, and Sibir were established. The Khanate of Kazan was annexed by Russia in 1552, whereby Tatars, Bashkirs, and Chuvash came under Russian rule. The West Siberian Tatars are partly descendants of Volga Tatars, who left their homeland in this period. A Tatar Autonomous Soviet Republic was established in 1920. After

the Soviet era, the Autonomous Republic of Tatarstan became a member of the Russian Federation.

Related Languages and Language Contacts

The Tatar language is related to Bashkir, Crimean Tatar, Kazakh, Karachay–Balkar, Kumyk, Karaim, etc. Tatar has influenced neighboring languages such as Bashkir, Chuvash, and the Finno-Ugric languages Mari (Cheremis), Mordva, and Udmurt (Votyak). The literary language has also had considerable influence on Turkic languages in Central Asia, e.g., Uzbek. Literary Tatar has to a certain extent served as a model for literary Kazakh. Tatar has been influenced by Russian, particularly in the lexicon. The written language was also used by the small Turkic groups of western Siberia, and thus had a strong impact on their dialects.

Certain features typical of Tatar are already found in the Kuman language as attested in the *Codex cumanicus* (14th century), where the language is even referred to as ‘Tatar.’ The written language used in the cultural centers of the Golden Horde was Khorezmian Turkic, which had its center in Khorezm on the shore of the Aral Sea and was influenced by local Kipchak and Oghuz Turkic dialects.

This tradition was continued in the Khanates that emerged after the fall of the Golden Horde. It was used, with strong Kipchak elements, as the official language in the Crimea up to the 17th century, when it was replaced by Ottoman (Turkish).

Its use in the Khanate of Kazan was strongly influenced by Chaghatay (Chagatai) and Ottoman. A so-called Volga Turki developed, which is often referred to as ‘Old Tatar,’ though it must be distinguished from older spoken Tatar. It was used for an emerging Tatar literature based on Chaghatay traditions. Religious works were written in this language up to the mid-19th century.

A more genuinely Tatar written language developed in the second part of the 19th century. It was based on the Kazan dialect, though strongly influenced by Chaghatay. It was also used by Mishers and Astrakhan Noghays and, for some decades, Bashkirs. It was of great cultural importance for all Turkic minorities in Russia. At the beginning of the 20th century, Tatar still had a considerable transregional validity. In the Soviet era, it was limited to a regional national language.

Tatar was written with Arabic script until a Roman-based alphabet was introduced in 1927. In 1939, a variant of the Cyrillic alphabet was adopted. The Christian Tatars in the Volga region had used the

Cyrillic script already at the end of the 19th century. In the post-Soviet era, a new Roman-based alphabet has been created, although it has not yet replaced the Cyrillic-based script.

Distinctive Features

Tatar exhibits most linguistic features typical of the Turkic family (see **Turkic Languages**). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation, e.g., *A = a/e*. A segment in round brackets only occurs after consonant-final stems. Hyphens are used here to indicate morpheme boundaries.

Phonology

The phonetic basis of modern standard Tatar is Kazan Tatar. The vowel system includes the high-mid vowels *ě, ö, ĭ, ǫ*, which are shorter and more centralized than the low and high vowels. The vowel *a* of the first syllable is rounded to *â* in the central dialect.

Tatar exhibits the results of systematic vowel shifts. Low vowels of the first syllable have been raised: *e > ĭ*, e.g., *min* ‘I’ (< *men*), *o > u*, e.g., *qul* ‘arm’ (< *qol*), *ö > ü*, e.g., *küz* ‘eye’ (< *köz*). High vowels have been centralized and reduced: *i > ĭ*, e.g., *bēr* ‘one,’ *u > ǫ*, e.g., *qōš* ‘bird’ (< *quš*), *ü > ǫ*, e.g., *kōn* ‘day’ (< *kün*).

Tatar suffixes display front vs. back harmony. The vowel of the suffix depends on the frontness vs. backness of the last stem syllable. The high suffix vowels are *ě* and *ĩ*, and the low suffix vowels are *e* and *a*, e.g., *ēt-ler-ěbēz-den* [dog-PL-POSS.1.PL-ABL] (front) ‘from our dogs’ vs. *at-lar-ībīz-dan* [horse-PL-POSS.1.PL-ABL] (back) ‘from our horses.’ Rounded vs. unrounded harmony is absent in Standard Tatar, which means that *ǫ, ǭ, u* and *ü* do not occur in suffixes.

Initial *ǰ*- is sometimes found instead of *y*-, mostly in front of *ĩ* and *i*, e.g., *ǰır* ‘place’ (cf. Turkish *yer*). In loans originating from Arabic (Arabic, Standard), *‘ayn* is represented by the voiced fricative *ɣ*, e.g., *ɣadet* ‘habit’ (< *‘a:dat*). The consonants that correspond to the affricates *č* and *ǰ* in most other Turkic languages (and are traditionally transcribed so) are pronounced as palatalized fricatives *ʃ* and *ʒ* in Standard Tatar. In front of suffix-initial vowels, stem-final *p*, *q*, and *k* mostly become *b*, *ɣ*, and *g*, respectively, e.g., *tab-a* [find-PRES] ‘finds’ vs. *tap* ‘find!’, *čig-ě* [boundary-POSS.3.SG] ‘its boundary’ vs. *čik* ‘boundary.’ Various assimilations affect suffix-initial consonants. Thus, the *l* of the plural suffix and the *d* of the ablative suffix are assimilated to *n* after stems

ending in nasal consonants, e.g., *uram-nar* [street-PL] ‘streets,’ *urman-nan* [forest-ABL] ‘from the forest.’ Nonpermissible consonant clusters are dissolved by means of epenthetic vowels, consonant deletion etc., e.g., *dus* ‘friend’ vs. *dust-ım* [friend-POSS.1.SG] ‘my friend.’

Grammar

After first- and second-person possessive pronouns the possessive suffix on the head is optional, e.g., *min-ëm eş-ëm* [I-GEN work-POSS.1.SG] or *min-ëm eş* [I-GEN work] ‘my work.’ The comparative degree of adjectives takes the suffix *-rAK*, e.g., *özön-raq* [long-COMP] ‘longer’ (cf. Turkish *daha uzun*). The third-person personal pronouns are *ul* ‘he, she, it’ (with the oblique stem *an-*) and *alar* ‘they.’ The demonstrative pronouns *bu*, *şuşı*, *şul*, *təğə*, *ul* express various degrees of proximity. Approximative numerals are formed with the suffix *-lAp*, e.g., *un-lap* ‘approximately ten.’

Tatar has numerous simple and compound aspect-mood-tense forms as well as verbal nouns, converbs, and participles. It has a present tense in *-A* (*-y* after stem-final vowels) plus personal markers, e.g., *kil-e-m[ən]* [come-PRES-1.SG] ‘I come, I am coming.’ The most frequent verbal noun ends in *-(I)w*, e.g., *al-ñw* ‘to take, taking.’ An infinitive is formed with *-(I)rGA*, negated *-mAs-kA*. Frequently used converb markers include *-A* (*-y* after stem-final vowels), *-(I)p-* and *-GAč*, e.g., *al-yač* [take-CONV] ‘after having taken.’ Like most other Turkic languages, Tatar has evidential markers of the type *iken*, e.g., *qayt-qan iken* [return-POSTTERMINAL.PAST EV] ‘has obviously returned.’ A number of auxiliary verb (postverb) constructions express modifications of the manner of action, e.g., *yan-ıp bēt-* [burn down-AUX] ‘burn down (completely).’ Possibility and impossibility are expressed by means of a converb + the auxiliary verb *al-*, e.g., *yaz-a al-* [write-CONV-POSS] ‘to be able to write,’ *yaz-a al-ma-* [write-CONV-POSS-NEG] ‘to be unable to write’ (Turkish *yaz-a-bil-* [write-CONV-POSS], *yaz-a-ma-* [write-CONV-POSS-NEG]).

Lexicon

Most basic lexical elements are of Turkic origin. Many loans are of Middle Mongolian, Arabic,

Persian, and Russian origin, e.g., *zur* ‘big,’ *aždaha* ‘dragon,’ *baqča* ‘garden,’ *atna* ‘week’ (Persian), *fikër* ‘thought,’ *taraf* ‘side’ (Arabic), *stakan* ‘glass,’ *par* ‘steam,’ *kuxnya* ‘kitchen,’ *vrač* ‘doctor’ (Russian), *uram* ‘street,’ *dala* ‘steppe’ (Mongolian). Words of Finno-Ugric origin occur mainly in dialects. Tatar conjunctions are mostly of foreign origin, e.g., *hem* ‘and,’ *emme* ‘but,’ *čönki* ‘for (causal),’ *güye* ‘as if,’ *ki* ‘that,’ *eger* ‘if, when.’

Dialects

Tatar comprises a central dialect group, Kazan Tatar proper. A western dialect group, consisting mainly of Mishar Tatar, is spoken in the Volga region outside the republic. An eastern dialect group is spoken in West Siberia. The Irtysh–Tobol dialects hold an intermediate position between Kazan Tatar and other Siberian Tatar dialects. West Siberian dialects often exhibit the changes *č > ds* and *ǰ > dz* and voicing of intervocalic consonants (like in South Siberia). The vowel shifts are not so strongly developed in these dialects as in Volga Tatar.

Bibliography

- Berta Á (1989). *Studia Uralo–Altaica 31: Lautgeschichte der tatarischen Dialekte*. Szeged: Universitas Szegediensis de Attila József nominata.
- Berta Á (1998). ‘Tatar and Bashkir.’ In Johanson & Csátó É Á (ed.) *The Turkic languages*. London/New York: Routledge. 283–300.
- Dawletschin T, Dawletschin I & Tezcan S (1989). *Tatarisch-deutsches Wörterbuch*. Wiesbaden: Harrassowitz.
- Johanson L (2001). ‘Tatar.’ In Garry J & Rubino C (eds.) *Facts about the world’s major languages: an encyclopedia of the world’s major languages, past and present*. New York: New England Publishing Associates/Dublin: The H. W. Wilson Company. 719–721.
- Poppe N (1963). *Tatar manual: descriptive grammar and texts with a Tatar–English glossary*. Indiana University Publications, Uralic and Altaic Series 25. Bloomington/The Hague: Mouton.
- Thomsen K (1959). ‘Das Kasantatarische und die westsibirischen Dialekte.’ In Deny J *et al.* (eds.) *Philologiae turicae fundamenta 1*. Aquis Mattiacis: Steiner. 407–421.

© 2006 Elsevier Ltd. All rights reserved.

Presented by: <https://jafrilibrary.com>

All the third-person pronouns are listed below, along with their oblique forms (that are explained later).

The 2pl and 3pl-nh pronouns are also used as honorific ('respect') pronouns. When honorificity is taken into account, in third person we get extra sets of pronouns that denote different degrees of 'respect.' The sets of pronouns with increasing degree of honorificity are 3s-mh REMOTE *wāḍu atanulāyana, wāru*; PROXIMATE *wīḍu, itanulāyana, wīru*; INTERROGATIVE: *ewaḍu, ewaru*; Third-person singular human feminine (as nonhumans are not differentiated for honorificity): REMOTE *adi, āmelāwīḍa, wāru*; PROXIMATE *idi, ime/wīḍa, wīru*; INTERROGATIVE *ēdi, ewarte, ewaru*. Note that the pronouns with highest degree of honorificity (*wāru, wīru, ewaru*) are originally one of the alternants of the Plural human forms (but the other alternants viz., *wāllu, wīllu, ewaḷu* are not used as third-person honorific singular forms).

Oblique Forms

Nouns (simple, derived as well as plural forms) and pronouns have both direct and oblique stems. The direct stems are the nominative forms (direct stems of singular nouns and ever-plurals (e.g., *pālu* 'milk') are listed in the lexicon), whereas the oblique stems are used in several of the case inflections. In the case of some nouns, both of these stems are same in form (e.g., *kukka* 'dog': *kukka-ki* 'to a dog'). In the case of all pronouns and some nouns, the oblique stems differ in shape (e.g., *rāyi* 'stone': *rāti-tō* 'with a stone'; *kukka-lu* 'dogs': *kukka-la-ki* 'to the dogs'). The oblique stems of third-person pronouns are given in Table 3. The oblique stems of the personal pronouns are *nā* (1s), *mā-* (1e-pl), *mana-* (1i-pl), *nī-* (2s), *mī-* (2pl).

Table 2 Vowels

| | <i>FU</i> | <i>CU</i> | <i>BR</i> |
|------|-----------|-----------|-----------|
| High | i ī | | u ūa |
| Mid | e ē | | o ō |
| Low | | a ā | |

Table 3 Third-person pronouns

| 3rd Person | Remote | | Proximate | | Interrogative | |
|--------------|-------------------|-------------------|-------------------|-------------------|--------------------|--------------------|
| | Direct stem | Oblique stem | Direct stem | Oblique stem | Direct stem | Oblique stem |
| <i>s-mh</i> | <i>wāḍu</i> | <i>wāḍi</i> | <i>wīḍu</i> | <i>wīḍi</i> | <i>ewaḍu</i> | <i>ewaḍi</i> |
| <i>s-nmh</i> | <i>adi</i> | <i>dāni</i> | <i>idi</i> | <i>dīni</i> | <i>ēd1</i> | <i>dēni</i> |
| <i>pl-h</i> | <i>wāru/wāllu</i> | <i>wāri/wālla</i> | <i>wīru/wīllu</i> | <i>wīri/wīlla</i> | <i>ewaru/ewaḷu</i> | <i>ewar1/ewaḷa</i> |
| <i>pl-nh</i> | <i>aw1</i> | <i>wāṭi</i> | <i>iwi</i> | <i>wīṭi</i> | <i>ēwi</i> | <i>wēṭi</i> |

Noun

A noun is simple (monomorphemic) (e.g., *kukka* 'dog') or derived (e.g., *donga* 'thief' > *donga-tanam* 'theft'; *moga-* 'male' > *moga-tanam* 'manliness'; *cuttam* 'a relative' > *cutta-rikam* 'relationship'; *andam* 'beauty' > *anda-gatte* 'beautiful woman'; *tiṇḍi* 'eating' > *tiṇḍi-pōtu* 'glutton'; *jūdam* 'gambling' > *jūdari* 'gambler'). Verbs can give rise to two types of derived nouns: action nominals (e.g., *mūyu* 'to close, cover' > *mūy-aḍam* 'closing, covering'; *pilucu* 'to call' > *pilawa-dam* 'calling') or substantive nominals (*mūyu* 'to close, cover' > *mū-ta* 'a cover, lid'; *pilucu* 'to call' > *pilu-pu* 'invitation').

Number

A simple or derived noun can be inflected for plural number by the addition of a plural suffix. The plural suffix has two morphophonemic alternants: *-lu* and *-ḷu*. [e.g., 'dog': *kukka* (sg.), *kukka-lu* (pl.); 'backyard': *peraḍu* (sg.), *peraiḷu* (pl.); 'cat': *pilli* (sg.), *pillulu* (pl.); 'house': *illu* (sg.), *iḷlu* (pl.); 'eye': *kannu* (sg.), *kaḷlu* (pl.)]. Some nouns require their oblique stems to receive the plural suffix [e.g., 'pit': *goyyi/gōyi* (sg.), *gōtu-lu* (pl.); 'horse': *gurram* (sg.), *gurrā-lu* (pl.)].

Case

The direct stem of a singular or a plural noun functions as a noun in nominative case. Other case forms of nouns are obtained by means of several case suffixes and postpositions. Some of them are accusative *-niḷ-nu*; dative *-kiḷ-ku*; instrumental/sociative: *-tō*; ablative *-nunci*; comparative *-kaṇṭe*; and locative *-lō*. The oblique stem of a noun functions as its genitive form (e.g., *nā* 'my', *rāti* 'of stone' [*rāyi* 'stone']). A few postpositions are: *kinda* 'below', *mīda* 'above', *lōpala* 'inside', *mundu* 'in front of, before', *tarawāta* 'after.'

Numerals

Structure of the numerals follows the general Dravidian pattern. Cardinal numerals 1000, 100, and 1 to 10 are mono-morphemic. They are: *okaṭi* '1', *reṇḍu* '2', *māḍu* '3', *nālugu* '4', *aidu* '5', *āru* '6', *ēḍu* '7',

enimidi ‘8,’ *tommidi* ‘9,’ *padī* ‘10,’ *nūru/wanda* ‘100,’ and *vēyi/veyyi* ‘1000.’ The formula for forming decades is: 2, 3 and so on, followed by 10 (e.g., *nalabhai* [4–10] ‘40’). The formula for series between decades (e.g., 41–49) is: numeral for decade followed by 1–9 (e.g., *nala-bhai-reṇḍu* [4-10-2] ‘forty-two’). Ordinals are derived from cardinals by suffixation of *-awa* (> *ō*) (e.g., *āru-awa* > *ārawa/ārō* ‘sixth’).

Adjectives and Adverbs

Adjectival forms that function solely as modifiers of nouns or other adjectives (and nothing else) are very few in the language; for example, *ara* ‘half,’ *pāwu* ‘a quarter,’ *ceri* ‘each.’ These adjectives can be followed only by a noun. The demonstrative and interrogative roots: *ā* ‘that,’ *ī* ‘this,’ *ē* ‘which,’ are also adjectives. They are used before nouns (e.g., *ā pilla* ‘that girl,’ *ē pilla* ‘which girl’). Their variants can take various suffixes to give rise to different forms (e.g., *akkāḍa* ‘there,’ *appuḍu* ‘then,’ *aṭu* ‘that side,’ *alāga* ‘in that manner,’ *awatala* ‘on that side,’ *anni* ‘that many,’ *anta* ‘that much’). Even the third-person pronouns can be viewed as derivatives of these forms.

Some adjectives are bound and require a suffix or a noun to follow it (e.g., *tella* ‘white’: *tella-wāḍu* ‘white man,’ *tella-ni/-ṭi maniṣi* ‘white person,’ *tella-gā* ‘whitish,’ *tella-na* ‘whiteness’). A large number of adjectives are derived from other forms such as nouns, adverbs, and verbs. A noun in genitive case always functions as an adjective (e.g., *nā pustakam* ‘my book,’ *rāṭi gōḍa* ‘stone wall’). Some examples of adjectives derived from adverbs are: *alāṭi maniṣi* ‘a person of that type’ [*alā(ga)* ‘in that manner’]; *rēpaṭi pani* ‘work of tomorrow’ (*rēpu* ‘tomorrow’). A majority of nouns function as modifiers when placed before another noun (e.g., *goppa maniṣi* ‘great man’).

It is difficult to find monomorphemic adverbs. Even the adverbs of time and place such as *ninna* ‘yesterday,’ *appuḍu* ‘then,’ are either basically nominals or are derived from adjectival roots. The main adverb deriving suffix is *-gā*, as in *gaṭṭi-gā* ‘hard.’ Many onomatopoeic words are basically adverbial in function (e.g., *gaba-gabā* ‘quickly’).

Verb

Like in many other Dravidian languages, verb in this language has the most complex structure. A fully inflected verb contains a verb stem followed by optional suffixes. The stem is simple, derived, or compound. A simple stem is composed of one verb root (e.g., *caccu* ‘to die’). A derived stem contains a verbal or nominal root followed by a derivative suffix (e.g.,

cam-pu ‘to kill,’ *cam-pincu* ‘to cause to kill’); *ūh-incu* ‘to imagine’ (from *ūha* ‘imagination’). A compound verb stem contains a main verb followed by one or more auxiliary verbs (e.g., *waṇḍu-konu* ‘to cook for oneself’ [*waṇḍu* ‘to cook’], *waṇḍu-kona-bōwu* ‘to be about to cook for oneself’).

A verb stem is inflected for tense/mode, which takes a further pronominal suffix (in the case of finite verbs). Most of the inflectional suffixes have two or more allomorphs. Some of the finite tense/mode forms are – Imperative: *tinu* [<*tinu-u*] ‘(You sg.) eat!’ *tinanḍi* [<*tinu-aṇḍi*] ‘(You pl.) eat!’ Negative imperative: *tinaku* [<*tinu-aku*] > ‘(You sg.) don’t eat!’ *tinakanḍi* [<*tinu-aku-aṇḍi*] > ‘(You pl.) don’t eat!’ Past: *pilicēnu* [<*pilucu-ē-nu*] ‘I called’; *pilicindi* [<*pilucu-in-di*] ‘It called.’ Nonpast/Habitual: *tiṇṭāḍu* [<*tinu-tā-ḍu*] ‘He will eat/He eats’; *tiṇṭāmu* [<*tinu-tā-mu*] ‘We will eat/We eat.’ Durative: *tiṇṭunnāḍu* [<*tinu-tunnā-ḍu*] ‘He is/was eating’. Nonpast Negative: *tinanu* [*tinu-a-nu*] ‘I will not eat’; *pilawaḍu* [<*pilucu-a-ḍu*] ‘He will not call’. Hortative: *tindām* [*tinu-dā-mu*] ‘Let us eat!’ Sample paradigms of three regular tenses inflected for all the persons (for the verb *tinu* ‘to eat’) follow:

| | Past | Nonpast | Nonpast Negative |
|-----------------|---------------------------|----------|---------------------|
| 1s | tinnānu | tiṇṭānu | tinanu |
| 1e-pl/ 1i-pl | tinnāmu | tiṇṭāmu | tinamu |
| 2s | tinnāwu | tiṇṭāwu | tinawu |
| 2pl | tinnāru | tiṇṭāru | tinaru |
| 3s-mb | tinnāḍu | tiṇṭāḍu | tinaḍu |
| 3s-nmb | tinnādi> tinnadi>tindi | tiṇṭundi | tinadu |
| 3pl-b | tinnāru | tiṇṭāru | tinaru |
| 3pl-nb | tinnāyi | tiṇṭāyi | tinawu |

Nonfinite verbs do not terminate in a Pronominal suffix. They form nonfinite or subordinate clauses in a sentence. The resulting forms have adverbial, adjectival, or nominal functions. Some of the nonfinite forms are obtained by a single suffix such as: Perfective (e.g., *cadiw-i* ‘having read’), Negative Perfective: *cadau-aka* ‘not having read’; Durative: *caduwwu-tū* ‘while reading’; Conditional: *cadiw-itē* ‘if one reads’; Concessive: *cadiw-inā* ‘even if one reads.’ Some Nonfinite forms are obtained by adding more than one suffix or auxiliary (e.g., *cadau-aka-pō-tē* ‘if one does not read’). Relative participle forms are adjectival in function – they are: Past: *cadiw-ina* ‘one who read, one which was read’; Nonpast: *cadiw-ē* ‘one who reads, one which is/will be read’; Negative: *cadau-ani* ‘one who does/did not read, one which is/was not read.’ The other important nonfinite forms are Verbal noun (e.g., *cadauḍam* ‘reading’), and Infinitive which

forms the basis for many further expansions (e.g., *cadawa*, as in *cadawa-kūdadu* ‘one should not read’).

Verbs are classified into different conjugation classes that account for the various morphophonemic changes that they undergo during the process of inflection.

An extensive process of verbal compounding gives rise to forms expressing different kinds of modes. These compound verbs can be classified on the basis of the inflected form of the nuclear verb. Some examples follow. *With infinitive as the nucleus*: Permissive: *cadawa-waccu* ‘one may read’; Inceptive: *cadawa-bō-yēnu* ‘I was about to read’; Potential: *cadawa-gala-nu* ‘I can read’; Negative Potential: *cadawa-lē-nu* ‘I cannot read’; Negative Past: *cadawa-lēdu* ‘One did not read’; Obligative: *cadaw-āli* ‘One should read’; Negative Injunctive: *cadawa-kūdadu* ‘One should not read’; Prohibitive: *cadawa-waddu* ‘Don’t read.’ *With past participle as the nucleus*: Benefactive: *cadiwi-peṭṭ-ēnu* ‘I read it (for somebody)’; Decisive: *cadiwi-tīru-tānu* ‘I will definitely read’; Completive: *cadiwi-wēs-ēnu* ‘I finished reading.’

Syntax

Telugu is an SOV language. “It is a nominative-accusative language and hence, the verb agrees with the argument in the nominative case. It has postpositions and the genitive precedes the governing noun. The comparative marker follows the standard of comparison. The complementizer occurs in the right peripheral position. Adjectives and participial adjectives precede the head noun. There are no pleonastic or expletive constructions such as *it* or *there*. It is a pro-drop language. The subject, direct object, indirect object, and adverbial phrase of the finite embedded and matrix sentence may be pro-dropped. There occur clefts in Telugu and the clefted constituent occurs as the rightmost element just as in other Dravidian, Tibeto-Burman languages and Sinhalese” (Subbarao and Bhaskararao, 2004: 161). It has four

kinds of nonnominative constructions and different types of verb-less sentences.

Vocabulary

Like many other Dravidian languages, the vocabulary of Telugu contains native Dravidian as well as borrowed vocabulary. The earliest borrowings were mostly from Sanskrit and Prakrit. Some of the borrowings were assimilated to fit the native phonology. Later borrowings came from Perso-Arabic sources through Urdu, Portuguese, and English in that order. Except for the sound [ʃ] all the other sounds of the borrowed sources that are not native to Telugu were replaced by nearer native sounds. Because verbal vocabulary is more resistant to accepting borrowals, although borrowing verbal concepts, the corresponding nouns from the source language were borrowed, which were verbalized by means of suffixation (e.g., *ūhincu* ‘to imagine’ [Sanskrit *ūha* ‘imagination’], *ānandincu* ‘to enjoy’ [Sanskrit *ānanda* ‘happiness’ or by means of verbal conjuncts (e.g., *ḍraywū-cēyu* ‘to drive’ [English *drive*]; *pūja-cēyu* ‘to worship’ [Sanskrit *pūjā* ‘workship’]).

Bibliography

- Bhaskararao P (1975). ‘Verbal compounding in Telugu.’ *Bulletin of Deccan College Research Institute* 35, 9–20.
- Bhaskararao P (1982). ‘A re-examination of consonantal Sandhi in modern colloquial Telugu.’ *Bulletin of Deccan College Research Institute* 41, 16–26.
- Krishnamurti Bh (1998). ‘Telugu.’ In Steever S B (ed.) *The Dravidian languages*. London: Routledge. 202–240.
- Krishnamurti Bh & Gwynn J P L (1985). *A grammar of modern Telugu*. Delhi: Oxford University Press.
- Subbarao K V & Bhaskararao P (2004). ‘Non-nominative Subjects in Telugu.’ In Bhaskararao P & Subbarao K V (eds.) *Non-nominative subjects*, vol. 2. Amsterdam: John Benjamins. 161–196.
- Subrahmanyam P (1974). *An Introduction to modern Telugu*. Annamalainagar: Annamalai University.

Thai

T J Hudak, Arizona State University, Tempe, AZ, USA

© 1994 Elsevier Ltd. All rights reserved.

Thai (Siamese, Central Thai) serves as the national language of Thailand where it is used by the schools, the media, and the government. Of the 1990 estimated population of 54 890 000, 75 percent are considered

ethnic Thai, 14 percent Chinese, and 11 percent other. Outside of Bangkok and the central plains, other regional dialects exist: Northern Thai (Kam Muang) in the north, Southern Thai in the south, and Lao or Northeastern Thai (Isan) in the northeast.

Thai belongs to the Tai language family, a subgroup of the Kadai or Kam-Tai family, and descended from the single protoparent Proto-Tai. A number of

Historical Background

In 1351, the Thai capital shifted to Ayutthaya. The most generally accepted theory holds that present-day Thai descended from the Sukhothai dialect. During the Ayutthaya period (1351–1767), Thai underwent two major changes. First, sometime between the mid-fourteenth and mid-seventeenth centuries, the system of three tones split into a system of five, the changes dependent upon the phonetic nature of the initial consonant of each syllable. Another significant change was the large influx of Sanskrit, Pali, and Khmer loanwords, which expanded the vocabulary and reflected the growing complexity of Ayutthayan society. Later, during the Bangkok era (1782–twentieth century), much of this terminology and its correct use became standardized by King Mongkut (1851–68). Further emphasis upon the correct use of the language came from King Chulalongkorn (1868–1910) and King Vajiravudh (1910–26). Since then,

Phonology

Syntax

Characteristic complex verbal predicates consist of a collocation of verbs referred to as serial verbs:

/pəy aw maa dāt plœœŋ kœœ khǎy tham sǎ mǎy/ 'She) went and got it changed it around, fixed it up, and made it just like new (Diller 1988: 280). These series often consist of a main verb modified by two sets of verbs, one preceding and the other following. Those verbs preceding often translate as English modals or adverbials, while those following often convey the sense of completion. In many cases the verbs are so arranged that they reflect the temporal sequence of the action.

Thai has three broad groups of particles that end utterances. One group marks a statement and forms questions that require yes–no answers; the second shows respect or deference toward the addressee; and the third indicates the mood of the speaker toward the situation at the time of speaking.

One of the most characteristic features of Thai is the use of classifiers, an obligatory class when quantifiers with nouns are present. The most usual order is noun + quantifier + classifier: /mǎa sǎam tual/ 'three dogs.' For each noun + classifier construction, the head noun determines the choice of classifier. Typical examples include /khon/ for human beings, /tual/ for animals, and /khan/ for vehicles and umbrellas.

Sociolinguistics

Beginning in the nineteenth century prior to the impact of Western languages, a type of traditional diglossia developed with the 'correct' speech based upon the speech of the royalty and upper classes (Diller 1988). Diller notes that much of this diglossia was characterized by vocabulary of Indic borrowings, although some syntactic patterns found in proper speech and formal written prose also appeared (Diller 1988: 304). With the impact of Western languages and the emphasis upon standardized grammars and languages in the nineteenth century, this diglossia became more and more solidified. A proliferation of

titles and ranks during the Ayutthaya period also helped to foster the idea of classes of speakers.

Another characteristic sociolinguistic feature of Thai is the complex pronoun system, with the choice of any one pronoun dependent upon factors such as age, sex, social position, and the attitude of the speaker toward the addressee. Pronouns are frequently omitted from surface syntax when the referent is understood. Kinship terms, and other nouns referring to relations, such as /phûan/ 'friend,' are often used as pronouns. Thus, /phôl/ 'father' may mean 'you, he' when speaking to or about one's father or 'I, father' when the father speaks to his child.

Future Work

Continued work on Thai will undoubtedly center upon the genetic relationship between Thai and other languages of southeastern and eastern Asia. A late twentieth-century controversy has revolved around the authenticity of the earliest known inscription, the Ramkhamhaeng Inscription.

Bibliography

- Chamberlain J R (ed.) (1991). *The Ram Khamhaeng Controversy: Collected Papers*. Bangkok: The Siam Society.
- Cooke J R (1968). *Pronominal Reference in Thai, Burmese, and Vietnamese*. Berkeley, CA: University of California Press.
- Diller A V N (1988). Thai syntax and 'National Grammar.' *Language Sciences* 10(2), 273–312.
- Gedney W J (1989). *Selected Papers on Comparative Tai Studies*. Michigan Papers on South and Southeast Asia, 29. Ann Arbor, MI: The University of Michigan.
- Hudak T J (1987). Thai. In Comrie B (ed.) *The World's Major Languages*. London: Croom Helm.
- Noss R B (1964). *Thai Reference Grammar*. Washington, DC: Foreign Service Institute.

Tibetan

P Denwood, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Tibetan comprises a multiplicity of regional spoken dialects, and a standardized written language (Classical Tibetan) which is the vehicle of a major civilization whose main religion is Buddhism. There are also several modern regional written languages.

Geography, Affiliation, and History

Tibetan is spoken in the Tibetan Autonomous Region of China, and in adjoining high-altitude parts of Bhutan, India (Ladakh in Kashmir and parts of Himachal Pradesh), Pakistan (Baltistan), Nepal (Mugu, Dolpo, Mustang, Solu Khumbu), Burma, and the Chinese provinces of Yunnan, Sichuan, Gansu, and Qinghai. Estimates of the number of speakers range from about three to seven million.

It is also used as a religious language by Mongols in the Republic of Mongolia, Inner Mongolia (China), and Russia (Buriats and Kalmucks), and by members of some ethnic groups in Nepal, including Newars and Tamangs, and other parts of the Himalayas. It is usually reckoned to be a member of the Tibeto-Burman language group, which, with the Karen and Chinese groups, forms the Sino-Tibetan family, though some scholars have cast doubt on this affiliation, citing parallels with Indo-European.

The Tibetans emerge into history in the 7th century AD. It is from that time also that their alphabetic writing system, based on a model of Indian origin, is alleged to date. The earliest datable example of the language is probably an inscription on a stone pillar in Lhasa dating from about 760 AD. Although originally often used for administrative purposes, since the 10th century Classical Tibetan has been closely associated with Buddhism, having been used to translate a vast range of literature, mostly from Sanskrit. There is also an indigenous literature, which was also almost entirely religious until the mid-20th century. Since that time the nonreligious genres of journalism and other 'nonfiction' have also flourished, and since the late 20th century also novels, short stories and poetry.

The spoken dialects have usually remained unwritten. Poorly recorded from premodern times, they have often developed separately from the written language and from one another. To ease the consequent difficulties of communication, several of the spoken dialects have come to be used as *lingua francas*: Lhasa Tibetan over the Tibetan Autonomous Region and among the exile community; Dzongkha in Bhutan, Leh Ladakhi in Ladakh and Amdo Khake (Amdo) in Qinghai and Kansu. While parallel modern regional written languages have also been developed (see below), the gap between spoken and written forms of the language remains wide.

Grammar

Words

A Tibetan word (phonologically defined) comprises a noun, verb, or adjective constituent, with or without one or more particles. A noun constituent may be polysyllabic, while verb and adjective constituents are all monosyllabic. Many verbs have variant forms ('stems' or 'roots') corresponding to tense/aspect differences. Other parts of speech are invariable, apart from sandhi variation with suffixed or prefixed particles. Particles express noun case categories and adjectival degree, mark the ends of subordinate clauses, and establish verb tense/mood/aspect categories. Most particles are suffixed, though a few

negative, dubitative, or interrogative ones are prefixed. There are also many phrasal nouns, verbs, and adjectives comprising two or more words. Particles are subdivided into noun, verb, and adjective, according to which type of word they occur in. A few particles can stand as separate words.

Noun Phrases

The order of elements in the noun phrase is: (1) head (noun), (2) epithet (adjective), (3) deictic (noun), (4) numerator (particle), (5) case marker (genitive, subject-marking, instrumental, dative-locative, ablative, comparative or adverbial particle).

Verb Phrases

In Classical Tibetan, a verb paradigm may have from one to four stems, and sometimes alternative forms for the same stem, e.g., the verb *seize* (shown here in transliterated Tibetan spelling):

| present | past | future | imperative |
|-------------|--------------|--------|------------|
| 'dzin./zin. | bzung./zung. | gzung. | zungs. |

A few verbs have suppletive paradigms with stems drawn from etymologically different verbs.

Modern spoken dialects show a reduction in the number and variety of verb stems; for example in the Lhasa dialect, for most speakers no verb has a separate future stem and many verbs have been reduced to a single stem: the equivalent of either the present or past of the classical language. More than compensating for this reduction has been a great increase in the use of verb particles and auxiliaries to express a complex mix of person, tense, aspect, mood, and evidential and judgmental modality systems. In many dialects there is also an unusual system of what may be termed 'viewpoint' – self-centered vs. other-centered – in which there is concord between the verb phrase and the speaker, who may or may not correspond to one of the arguments of the clause.

Verbs may be divided into two types: verbs of being (also used as auxiliaries) and lexical verbs. Verbs of being participate obligatorily in the grammatical systems of viewpoint and evidential modality. Lexical verbs are of two types, which determine their participation in the systems mentioned: 'intentional,' where the action is under voluntary control, and 'unintentional.' The majority pattern of the lexical verb phrase is: (1) lexical verb stem, (2) linking particle, (3) polar particle, (4) auxiliary, (5) modal particle.

Past, present and future tenses are established by a combination of verb stem and auxiliary. Similar means are used to distinguish perfect, progressive,

and prospective aspect, of which there are several subtypes in each case.

Clauses

Tibetan clauses are of SOV (subject-object-verb) type, with OSV order also possible. As well as the clause-final verb phrase, the clause may contain a subject, an object, and one or more adjuncts, all noun phrases. Subject and object phrases are regularly omitted without being represented by pronouns if they are not 'new.'

The main clause is the last in the sentence. Nonfinal (subordinate) clauses are usually marked by special particles.

Past-tense and often present-tense clauses are syntactically ergative, the subject of a transitive clause being marked with a particle identical in written form to the 'instrumental' noun particle.

Phonology

Modern central, southern, and eastern dialects have well-developed lexical tone, which has been analyzed in various ways, the simplest being as a two-tone system. In the Lhasa dialect the word is the domain of tone, which is manifested mainly as the pitch (high or low) of its first (or only) syllable. These tonal dialects mostly have few word-initial consonants, with few or no consonant clusters at word-initial position. Plosive and affricate initials always maintain a clear differentiation between an aspirated and an unaspirated series, while voicing has tended to disappear from these series as well as from fricative initials. The dialect of Dingri in southern Tibet has 27 word-initial consonants, all of them simple.

In the western dialects of Balti (spoken in northern Pakistan) and Ladakhi, as well as in some northeastern dialects of Gansu and Qinghai, tone is usually less well developed or absent, with a richer variety of word-initial consonant clusters. The northeastern Amdo Khake (Amdo) dialect has 36 simple word-initial consonants and 78 cluster initials. The writing system, whose spellings are full of consonant clusters, would suggest that the dialect it was based on, perhaps a central dialect of the 7th century, may have been pronounced somewhat like these so-called 'archaic' dialects. However, none of the present dialects approaches the complexity of the spelling system in this respect.

The tonal dialects of central and southern Tibet generally also have a system of vowel harmony. In the Lhasa dialect its domain is a pair of adjacent syllables within a word. Many noun, verb, adjective, and particle constituents vary between an open and a

close alternant. Most of the nonparticle constituents in question are spelt with one of the vowels *o*, *e*, or *a*: they will be pronounced with a closer vowel alternant when next to a syllable spelt with *i*, *u*, or the combination *ab*. There is little or no evidence of vowel harmony in the script, suggesting that its development, like that of tone, may have accompanied the progressive loss of consonant distinctions. Whereas the 'archaic' or 'cluster' dialects may typically have nine vowels, corresponding to the five of the script, the 'modern' or 'noncluster' harmonic dialects may have about 25 (in both cases, analyzed nonphonemically).

Honorifics

The written language and most of the dialects have a well-developed honorific system, in which lexical choice of verb is determined by the social status of the person acting as its grammatical subject. There is also a 'respectful' system, in which there is concord between choice of verb and direct or indirect object, and the two systems may be combined. Nouns, adjectives, and verb particles are also affected.

Sample Sentence

(Lhasa dialect: transliterated spelling in italics with phonetic rendering below: tones unmarked)

| | |
|--|----------------------------|
| <i>“a.las.</i> | <i>rang.gis.</i> |
| <i>ʔale:.</i> | <i>rangi</i> |
| EXCLAMATION | you-ERG SUBJ-MAKING PART |
| <i>zer.yag.la.</i> | |
| <i>sejala</i> | |
| say-NOMINALIZING PART-DATIVE-LOCATIVE PART | |
| <i>cha.</i> | <i>bzhag.na/ lha.sa'i.</i> |
| <i>təa</i> | <i>əaənə, ʔese:</i> |
| belief | place-if Lhasa-of |
| <i>gnam. gzhi.ni.</i> | <i>dgun.ka.</i> |
| <i>naməni</i> | <i>gyŋgə</i> |
| climate-TOPIC-MARKING PART | winter |
| <i>dro.po.</i> | <i>dang. dbyar.ka.</i> |
| <i>trəpo</i> | <i>tā jaagə</i> |
| warm-ADJ PART | and summer |
| <i>gsil.po.</i> | <i>yod.pa.'dra/</i> |
| <i>siibu</i> | <i>jə:bədra.</i> |
| cool-ADJ PART | is-seem |

'Well! To believe what you say, the climate of Lhasa seems to be warm in winter and cool in summer!'

Recent History

Developments since World War II have led to the political fragmentation of the Tibetan-speaking

world and the increasing influence of other languages, particularly Chinese (Mandarin Chinese), English, Urdu, Hindi, and Nepali. However, the same period has also seen the development of Modern Literary Tibetan (in Tibet and among refugees), Written Dzongkha (in Bhutan), and Written Ladakhi (in Kashmir) as written languages, based respectively on Lhasa Tibetan, spoken Dzongkha, and spoken Ladakhi, but influenced by Classical Tibetan. Some other dialects, including Amdo Khake (Amdo), Kham, and Sikkimese have also had written equivalents devised for them. The late 20th century has also witnessed a Tibetan diaspora, which has led to vastly increased interest in the language and culture, centered on a

numerically small but culturally active exile community in India and Nepal. Since the late 20th century, there has also been a marked revival of Tibetan in the Republic of Mongolia. Despite the problems experienced by its speakers, Tibetan remains a living, vigorous, and developing language.

Bibliography

- Denwood P (1999). *Tibetan*. Amsterdam/Philadelphia: John Benjamins.
Goldstein M C (1973). *Modern Literary Tibetan*. Urbana: University of Illinois Press.

Tigrinya

D Appleyard, University of London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Tigrinya (self-name *tigrinṃṃa* or *tigraj*), which is spoken in Eritrea and Ethiopia, is the second largest member of the Ethiopian branch of the Semitic family of languages, constituting together with Tigre and the extinct Ge'ez (or Classical Ethiopic) the northern subdivision. Estimates of the number of speakers in both countries vary from 4 to 5 million. Tigrinya is one of the two working languages of Eritrea, where it is the first language of about 50% of the population, and a major national language of Ethiopia. Tigrinya is written in a slightly expanded version of the Ethiopic syllabary, and as a written language has a history only from the latter half of the 19th century, due in great part both to the prestige of Ge'ez as the written language of Christian Ethiopia in the past, as well as to the dominance of its sister language, Amharic, as the language of the Ethiopian court.

Modern Tigrinya shows a considerable degree of dialect variation in the handful of preliminary studies that have been done. The standardization of written Tigrinya took its impetus from the full independence of Eritrea in 1993 and the adoption of Tigrinya as the principal language of the state.

Phonology

Tigrinya has 32 consonant and 7 vowel phonemes. Distinctive are the glottalized consonants and the labialized velars. The velars /k/, /kʷ/, /kʰ/ and /kʷʰ/ have fricative allophones in postvocalic position

including across close juncture between word boundaries: *kəfətə* 'he opened' but *jixəffit* 'he opens', *k'orbot* 'leather' but *ʔita x'orbot* 'the leather'. As the script has special symbols for these allophones, they will be indicated in the data here. Consonant length is phonemic, except for the glottals and pharyngeals which do not have lengthened counterparts. (See Table 1 for the consonant chart.)

Ethiopianist convention occasionally employs different symbols from the IPA ones used here; thus, *š* = *ʃ*, *ž* = *ʒ*, *č* = *tʃ*, *q* = *kʰ*, *ṭ* = *tʰ*, *č* = *tʃʰ*, *ǧ* = *dʒ*, *š* = *sʰ*, *p* = *pʰ*, *ṇ* = *m*, *y* = *j*, *k* = *x*, *q̣* = *xʰ*, *h* = *h*, *ʿ* = *ʕ*, *ʾ* = *ʔ*, *ä* = *e*, *ə* = *i*. The vowel phonemes of Tigrinya are */i/, /ɪ/, /u/, /e/, /o/, /ə/, and /a/*, of which the central vowels */ə/, /ɪ/, and /a/* are of particularly frequent occurrence. The mid-central vowel */ə/* has a markedly more open allophone in word final position: *nəgərə* = [nəgərə̃] 'he spoke', and indeed following a glottal or pharyngeal consonant this is written in the script with the same vowel sign as */e/*.

Morphology

Tigrinya, like other Ethiopian Semitic languages, has a complex inflectional morphology, particularly in the verbal system, employing not only prefixes and suffixes but also internal modification of the typical Semitic consonantal root-and-pattern type. Internal modification is also employed in forming many noun plurals from the singular, sometimes in combination with the addition of an affix, in 'broken plural' patterns so typical of other Semitic languages such as Arabic: *wərhi* 'month', *ʔawarhi* 'months', *kənfer* 'lip', *kənafir* 'lips'. Other noun plurals are formed

Table 1 The consonant phonemes of Tigrinya

| | <i>Bilabial</i> | <i>Alveolar/dental</i> | <i>Palatal</i> | <i>Velar</i> | <i>Pharyngeal</i> | <i>Glottal</i> |
|----------------------------------|-----------------|------------------------|----------------|--|-------------------|----------------|
| Plosive/affricate Glottalized | b p | d t | dʒ tʃ | g k | | ʔ |
| Plosive/affricate/fricative | pʼ | tʼ | tʃʼ | kʼ | | |
| Labialized | | | sʼ | [xʼ] g ^w k ^w kʼ ^w | | |
| | | | [xʷ][xʷ] | | | |
| Fricative | f | z s | ʃ ʒ | [x] | ħ h | h |
| Nasal | m | n | ɲ | | | |
| Lateral | | l | | | | |
| | | r | | | | |
| Approximant | w | | j | | | |

by suffixes alone: səb ‘man’, səbat ‘men’, g^wasa ‘shepherd’, g^wasot ‘shepherds’. Plural formations are determined lexically and cannot be predicted from the shape of the singular form.

In addition to number, nouns also have the category of gender, with two terms linked with male and female in animate nouns, while inanimates generally fluctuate in gender. Gender is mostly observable only in agreement. Definiteness is also indicated in the noun phrase by means of an article, in origin a remote demonstrative: ʔitom kahnat ‘the priests’, ʔita waʔro ‘the lioness’, ʔitu sʼibbuxʼ wəddi ‘the good boy’. Case relations are expressed by prepositions. Particularly interesting is the use of nɪ-(nɛ- + DEF) as optional marker of a definite direct object, the same clitic also having the function of indicating an indirect object:

nɛ-tu sirnaj ji-xirkiriʔ-o
 OBJ-DEF wheat 3(PL)-grind.IMPERF.PL-IT
 ‘they grind the wheat’
 ʔitu kʼəʃfi nɛ-ta səbəjti ji-ngrɛr
 DEF priest to-DEF woman 3MASC.SING-
 tell.JUSSIVE
 ‘let the priest tell [it] to the woman’

Verbs inflect for voice or valency, tense-mood-aspect (TMA), and person. In addition to the base-stem shapes, which are essentially three in number, there are three prefixed stem derivatives for each of these: tɛ-, which essentially has the function of marking passive-reflexive, ʔa- which generally has the function of marking causative-transitive, and a complex formative comprising ʔa-combined with lengthening of the first radical consonant or -t (i.e., formative ʔat-) before a glottal or pharyngeal. The meanings of derived stems are in addition often lexically defined. There are also specific TMA stem patterns associated with each of these derivational elements: nɛgrɛr ‘he spoke’, tɛ-nɛgrɛr ‘it was spoken’, bɛxɛjɛr ‘he wept’, ʔa-bkɛjɛr ‘he made someone weep’, lasʼɛjɛr ‘he shaved himself’, ʔa-lasʼɛjɛr ‘he made someone

shave himself’, fənnɛwɛr ‘he sent away’, ʔaf-fanɛwɛr ‘he accompanied someone on his way’, etc.

There are four fundamental TMA forms, conventionally referred to as the Perfect, the Imperfect, the Jussive-Imperative, and the Gerundive. The latter is sometimes also described as a Converb. These are marked both by different stem shapes and by different person markers, with the Imperfect and the Jussive having the same set of person markers (the Imperative marks only gender-number). The tense system is considerably augmented beyond these basic forms by means of auxiliaries and periphrastic constructions.

ʔab ʔasmɛra tɛ-wɛləd-ku
 in Asmara PASS-bear.PERF-1SING.PERF
 ‘I was born in Asmara’

maj ʔa-ʔilliħ
 water 1SING.IMPERF-CAUS-boil.IMPERF
 ‘I boil the water’

nɛ-tu gənzeb ki-[i]-ħib-ekka ʔij-jɛ
 OBJ-DEF money FUT-[1SING.IMPF]-
 give.IMPERF-you COP-1SING
 ‘I will give you the money’

tɪ-ħiz-o ʔallo-xa
 2MS.IMPERF-catch. be.PRES-2SING.MASC[PERF]
 IMPERF-him
 ‘you catch him (now)’

ʔab-zu ʔabij gɛza ji-xʼimmɛtʼ nɛbɛr-ɛ
 in-this big house 3MASC.SING. be.PAST-3MS.
 IMPERF-live. PERF
 IMPERF
 ‘he was living in this big house’

sʼibah tɛmɛlis-ɛ ʔi-xɛwwin
 tomorrow return.GER- 1SING.IMPERF-be.IMPERF
 1SING.GER
 ‘I may come back tomorrow’

The gerundive is used both as a subordinate verb, marking an anterior event in a sequence, and as a main verb form, expressing the result of an action:

mis men mes'iʔ-ka
 with who come.GER-2MASC.SING.GER
 'with whom have you come?'

nab-tu geza temelis-a
 into-DEF house return.GER-3FEM.SING.GER
 tɛx'ɛmmit'-a ʔingera
 sit.GER-3FEM.SING.GER bread
 hab-ɛtt-o
 give.PERF-3FEM.SING.PERF-him
 'she returned to the house, sat down, and gave him
 some bread'

Syntax

Word order in Tigrinya is generally subject-object-verb (SOV), with subordinate clauses preceding the main clause. Noun phrases are also generally head final with modifiers, including relative clauses, preceding the noun.

| | |
|--------------------------|--------------|
| ʔitu ʔanbesa zi-x'etel-ɛ | seb bi-h |
| | ak'k'i |
| DEF lion REL-kill.PERF- | man in-truth |
| 3MASC.SING.PERF | |

zi-Ø-ferrih
 REL-[3MASC.SING.IMPERF]-fear.IMPERF
 ʔaj-kon-ɛ-n
 NEG-be.PERF-3MASC.SING.PERF-NEG
 'the man who has killed a lion indeed has nothing to
 fear'

Bibliography

- Kane T L (2000). *Tigrinya-English dictionary*. Springfield: Dunwoody Press.
- Kiros F W (1985). *The perception and production of Tigrinya stops*. Uppsala: Uppsala University Department of Linguistics.
- Kogan L E (1997). 'Tigrinya.' In Hetzron R (ed.) *The Semitic languages*. London & New York: Routledge. 424–445.
- Leslau W (1941). *Documents tigrigna (éthiopien septentrional): Grammaire et textes*. Paris: Librairie C. Klincksieck.
- Tesfaye T Y (2002). *A modern grammar of Tigrinya*. Rome: Tipografia U. Detti.
- Voigt R (1987). *Das tigrinische Verbalsystem*. Berlin: Reimer Verlag.
- Ullendorff E (1985). *A Tigrinya (Təgrāñña) chrestomathy. (Äthiopistische Forschungen 19)*. Stuttgart: Franz Steiner Verlag.

Tiwi

J R Lee, Summer Institute of Linguistics, Darwin, NT, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Tiwi is an Australian Aboriginal language spoken by the Tiwi people, who number about 2000 and live on Bathurst Island and Melville Island (north of Darwin). Over the decades, since the first extensive contact with Europeans early in the 20th century, the Tiwi culture has undergone considerable change; the Tiwi people have changed from a seminomadic, hunter and gatherer way of life to a more settled lifestyle, and now mainly live in four townships on both islands. The Tiwi people are caught between two cultures, traditional and modern; they desire the benefits of European culture but also want to retain their own identity through some of their traditional ways. Although Tiwi people still do some hunting and gathering and maintain some of their traditional ceremonial life, they are now dependent on a money economy and are mostly Roman Catholic in religion. This change is also reflected in what has happened and is still happening in the language.

The language change is so extensive that young people (even people in their forties) no longer speak or even understand much of the traditional language. Not only does the language of the young people contain a number of English words, but the actual structure of the language has changed. These changes are due to a combination of factors over a period of decades. One of the most significant factors was the setting up of a school for both boys and girls in 1914, with English as the language of instruction and literacy. In addition to this, between 1921 and 1973, most girls were brought up from about the age of six in a dormitory, which effectively cut them off from extensive contact with their families and from hearing the Tiwi language spoken in a regular family context.

The verbal repertoire of the Tiwi people can be characterized by at least five codes: Traditional Tiwi, Modern Tiwi (a modified form of Traditional Tiwi), New Tiwi (an anglicized Tiwi), Tiwi-English, and Standard Australian English. These codes, though having characteristics that distinguish them from each other, are not discrete, but rather merge into one another along a spectrum. Each code has within it characteristic styles. For instance, within New Tiwi, there is a difference between the more formal style,

used in storytelling on tape and in elicited speech, and the less formal style, used in spontaneous speech. Also, the New Tiwi used by children is different from that used by adults. The Tiwi code used by a person is largely dependent on the age of the speaker, but not exclusively so. Though most young people do not command much Traditional Tiwi (with their understanding being greater than their production), older people do appear to command New Tiwi to some extent and usually use it in speaking with younger people. In addition to the diversity of codes, the situation is made more complex by switching between codes, including English.

Traditional Tiwi

In Traditional Tiwi (TT), there are four vowels: a, i, o, u. The consonants are given **Table 1**, in which the symbols used are the orthographic ones developed when Tiwi became a written language 30 years ago. The prenasalized and labialized stops are interpreted as single consonants on the basis that there are no unambiguous double consonants in Tiwi. The Tiwi syllable pattern is consonant-vowel (CV) or V with no closed syllables.

Tiwi nouns are divided into two classes, masculine and feminine. For humans (and some animals), the distinction is on the basis of natural sex. For nonhumans, the distinction is made on other criteria, normally semantic grounds. Plurality is marked only on human nouns, and the distinction between masculine (MASC) and feminine (FEM) is lost in plural (PL) nouns. Adjectives agree with the nouns that they qualify in gender and plurality (when applicable), as shown in the following examples:

- (1) arikula-ni tini
 big-MASC man
 ‘big men’
 arikula-nga tinga
 big-FEM woman
 ‘big women’

arikula-pi tiwi
 big-PL people
 ‘big people’

Traditional Tiwi is a polysynthetic language, with the inflected verb having an extremely complex structure. It is one of the prefixing languages of northwestern Australia but it has not been found to be directly related to any of them. The verb is able to take a number of affixes (mainly prefixes), indicating subject person, direct or indirect object person, tense, aspect, mood, time of day, and distance in time or space, for example. The nucleus of the verb contains a verb root but may also contain one or more incorporated forms that add some other nominal, stative, or verbal meaning (abbreviations: CONT, continuous; INCL, inclusive; SUBJUNC, subjunctive; EMPH, emphatic; CON, connective; CAUS, causative):

- (2) Pi-rri-mini-wujingi-pirni.
 they-PAST-me-CONT-hit
 ‘They were hitting me.’
 (3) Yinkiti nga-ma-wun-ta-y-akirayi.
 food we(INCL)-SUBJUNC-them-EMPH-CON-give
 ‘We should give them food.’
 (4) Nganti-ri-ma-rri-pi-y-ajirringi-kitikim-ani warta.
 we.PAST.FEM-CON-with-CON-bush
 long.thing-CON-crocodile-drag-PAST.HABIT
 ‘We used to drag the crocodile to the bush with
 the spear still in her.’
 (5) Taringini yi-mini-maji-wutu-wirri.
 snake he.PAST-me-on-horse-bite.
 ‘A snake bit me while I was on a horse.’

In Traditional Tiwi, there is also a verb phrase, consisting of a free-form verb, which carries the basic meaning, and an auxiliary verb, which can carry the same inflections as an independent verb. The class of free-form verbs occurring in this type of construction is small and, even in TT, may be expanded by the use of English loan verbs.

Table 1 Traditional Tiwi consonants

| Feature | Apical | | Laminal | Peripheral | |
|--------------------|----------|--------------|----------------|----------------|--------|
| | Alveolar | Postalveolar | Dental/Palatal | Dorsal | Labial |
| Stops | t | rt | j | k | p |
| Prenasalized stops | nt | rnt | nj | nk | mp |
| Nasals | n | rn | ny | ng | m |
| Labialized stops | | | | kw | pw |
| Labialized nasals | | | | ngw | mw |
| Laterals | l | rl | | | |
| Rhotics | rr | r | | | |
| Semivowels | | | y | g ^a | w |

^aThe symbol ‘g’ represents a velar fricative, which seems to behave like a semivowel.

- (6) Papi awungarra pi-ri-maji-wutuwu-mi.
arrive here they.PAST-CON-on-horse-do
'They arrived here on horses.'
- (7) mwarliki nga-ma-wun-ta-m-amigi
bathe we(INCL)-SUBJUNC-them-EMPH-do-CAUS
'we should cause them to bathe'

New Tiwi

The speech of young people incorporates a number of changes to the traditional language, including phonological changes and changes in vocabulary, in noun classification, and in syntax, such as word order. However, the greatest change is in the verbs. New Tiwi (NT) is no longer a polysynthetic language but has become more isolating. Most of the verbal inflection has been lost, with simple verb forms mostly replacing the complex inflected verbs. The NT verb form is based on the TT verb phrase. However, the small class of free-form verbs has been expanded by a greater use of loan verbs from English and, in a few cases, by the use of the singular imperative as a free-form verb. The auxiliary verb may or may not be used, depending on the formality of the occasion. When it is used, there are very few inflections retained, usually only those prefixes marking subject and tense, though even these are often changed in form. The following three examples are comparisons of New Tiwi and Traditional Tiwi:

- (8a) NT: Wokapat yi-mi.
walk he.PAST-do
'He walked.'
- (8b) TT: Yi-p-angurlimayi.
he.PAST-CON-walk
'He walked.'
- (9a) NT: Lukim ngi-ri-mi nginja.
see I-CON-do you.SING
'I saw you.'
- (9b) TT: Ngi-rri-min-j-akurluwunyi (nginja).
I-PAST-you(SING)-CON-see (you.SING)
'I saw you.'
- (10a) NT: Tamu ji-mi.
sit she.PAST-do
'She sat.'
- (10b) TT: Ji-yi-muwu.
she.PAST-CON-sit
'She sat.'

Note: *tamuwu* is the singular imperative form in Traditional Tiwi.

Young people sometimes use the continuous action prefix *wuji-*, but other aspects and moods are given by loanwords from English, such as *stat* 'start,' *tra* 'try,' *jut* or *shut* 'should,' and *ken* 'can.'

- (11a) NT: Yoyi a-wuji-ki-mi.
dance he-CONT-CON-do
'He is dancing.'
- (11b) TT: Yoyi a-wuji-ngi-mi.
dance he.NONPAST-CONT-CON-do
'He is dancing.'
- (12) NT: Jirra tra kirrim ji-mi warra.
she try get she.PAST-do water
'She tried to get some water.'

Since the verbs in NT no longer have the inflections of TT, there is a greater use of nouns and pronouns, to indicate the participants, and dependence on time words or the context, to indicate the tense. In NT, there is also considerable use of other English loanwords. In the speech of young people, particularly children, these may include words such as *pijipiji* for 'fish,' for which there is a traditional Tiwi equivalent. In general, NT can be said to be an 'amalgam' of Tiwi and English – in other words, a 'mixed code,' like a pidgin (or creole) Tiwi. This type of amalgam distinguishes 'mixing' from 'switching' between codes, though it is often hard to tell where mixing ends and switching begins. The following example from a 4-year-old boy shows the mixing in his Tiwi and the switching between his New Tiwi and his Tiwi-English (TE) in the same utterance:

- (13) Ya kilim ja. [NT] I kill you, mate. (TE)
I hit you (SING) I hit you mate
'I'll hit you, mate.'

Modern Tiwi

What is normally thought of as Modern Tiwi (MT) is a style, or a range of styles, between Traditional Tiwi and New Tiwi; in general, Modern Tiwi is a modified or simplified style of TT. In Modern Tiwi, people use more verbs with TT verb roots than are used in NT, but the verbs do not have the same richness of inflection as in TT. In general, the only affixes retained in the verb are the subject and tense prefixes and others that are not able to be expressed externally in TT, such as some aspect and mood affixes. Other affixes and incorporated forms are normally omitted, particularly those affixes indicating whether an action was done in the morning or evening and the object prefixes. These are either expressed by free-form words or are understood from the context.

- (14a) MT: Japinari yi-pirni ngiya.
morning he.PAST-hit me
'He hit me in the morning.'
- (14b) TT: (Japinari) yu-watu-mini-pirni.
morning he.PAST-morning-me-hit
'He hit me in the morning.'

Another feature of MT is the loss of distinction between first-person plural inclusive and exclusive subjects, as is also the case in NT.

- (15a) TT: Ngimpi-ri-majirripi.
 we(EXCL)NONPAST-CON-lie.down
 'We (but not you) lie down.'
 Nga-ri-majirripi.
 we(INCL)-CON-lie.down
 'We (including you) lie down.'

- (15b) MT: Nga-ri-majirripi.
 we.NONPAST-CON-lie.down
 'We lie down.'

Conclusion

The present-day language situation among the Tiwi people is very complex; a broad overview cannot begin to describe the differences that there may be among the four townships on Bathurst and Melville islands. Briefly, there are certain domains wherein a particular code is appropriate and may be used exclusively (or almost exclusively). In other situations, more than one code may be used, depending on the speakers and hearers and the formality of the occasion. Traditional Tiwi is used in the traditional ceremonies and songs, in some liturgy in the church, and in some written material. In situations in which non-Tiwi people are involved, such as administrative and work situations, English is mostly used, though the English used by the Tiwis may vary between Standard English and Tiwi-English. English is also used in the homily and in some of the liturgy and songs in church, and in all of the schools, at least in formal education. Modern Tiwi is used in bilingual education in the primary school on Bathurst Island and in some formal situations, when young Tiwi people are involved,

such as when giving a formal speech in Tiwi. There is some written material in Modern Tiwi, produced by the schools and by the author (Jennifer Lee) and her Summer Institute of Linguistics colleague, Marie Godfrey (New Tiwi is not generally acceptable in written form). New Tiwi is the common code of young people, though with considerable code switching with English in most situations, depending on the speakers and hearers.

Bibliography

- Godfrey M (1979). 'Notes on paragraph division in Tiwi.' In *Work papers SIL-AAB A-3*. Darwin: Summer Institute of Linguistics. 1-22.
- Godfrey M (1985). 'Repetition in Tiwi at clause level.' In *Work papers of SIL-AAB A-9*. Darwin: Summer Institute of Linguistics. 1-38.
- Godfrey M P (1997). 'Logical propositions in modern Tiwi.' In McLellan M (ed.) *Studies in Aboriginal grammars, SIL-AAIB occasional papers, 3*. Darwin: Summer Institute of Linguistics. 25-55.
- Lee J R (1987). *Tiwi today: a study of language change in a contact situation. Pacific linguistics series C no. 96*. Canberra: Research School of Pacific Linguistics, The Australian National University.
- Lee J R (1988). 'Tiwi: a language struggling to survive.' In Ray M J (ed.) *Aboriginal language use in the Northern Territory: 5 reports, work papers of the Summer Institute of Linguistics, Australian Aborigines and Islanders Branch B, 13*. Darwin: Summer Institute of Linguistics. 75-96.
- Lee J R (compiler) (1993). *Ngawurranungurumagi nginin-gawila ngapangiraga. Tiwi-English dictionary*. Darwin: Summer Institute of Linguistics, Australian Aborigines and Islanders Branch/Nguiu, Bathurst Island, NT: Nguiu Nginginawila Literature Production Centre.
- Osborne C (1974). *The Tiwi language*. Canberra: AAIS.

Tocharian

R Kim, University of Pennsylvania, Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Tocharian (Tokharian) is the conventional name for two related, extinct Indo-European (IE) languages known from documents found in the oases north of the Taklamakan desert in Xinjiang (Chinese Turkestan). The languages are now generally referred to as Tocharian A (TA) and Tocharian B (TB); the alternatives *Osttocharisch* and *Westtocharisch* (East

and West Tocharian) are still used by German scholars; 'Turfanian' and 'Kuchean' are obsolete terms.

Tocharian is known from manuscripts discovered by archaeological missions to Xinjiang in the years preceding World War I, in particular those led by Sir Aurel Stein of the United Kingdom, Albert von le Coq of Germany, and Paul Pelliot of France. In addition to a wealth of Middle Iranian documents, the expeditions brought back others in unknown languages, written in the 'slanting' Brāhmī script of Central Asia. In 1908 the German philologists Emil Sieg and Wilhelm Siegling conclusively identified them

as non-Indo-Iranian IE languages, which they labeled 'Indo-Scythian'; they also succeeded in distinguishing TA and TB.

The manuscripts are dated to approximately the sixth to eighth centuries AD, but further chronological precision is difficult. The TA records were discovered in and around Turfan and Qarašahr and are entirely of Buddhist religious content; most are translations or adaptations of Sanskrit originals. TB documents were found across the northern Silk Road from Kuča in the west to Turfan in the east; most are Buddhist in content, but a solitary love poem and a large number of monastery records, as well as caravan passes and cave graffiti, indicate that TB was the vernacular of at least part of the population in these areas in the later first millennium.

TA is remarkably uniform linguistically, and a number of facts indicate that it was no longer spoken at the time of the surviving manuscripts, but served as a sort of liturgical language among speakers of TB and Old Turkic. The TB documents exhibit considerable variation on all levels. On the basis of certain phonological and morphological features, they have been divided into western, central, and eastern dialects, but as vernacular TB sources (e.g., caravan passes or cave graffiti) mostly show 'eastern' characteristics, this division may also reflect chronological and/or sociolinguistic differences. Another source of variation is poetic: many forms in TB verse passages have been adjusted by one syllable to fit the meter; also characteristic is *pudñäkte* 'Buddha' for prose *pañäkte*.

The speakers of Tocharian played an important role in the Buddhist civilization of pre-Islamic eastern Central Asia, but their exact identity remains unknown. The name 'Tocharian' rests mainly on the form *twyry* in an Old Uyghur colophon, but both the reading and the identification have been challenged. It seems certain that the speakers of TA and TB were not the 'Tocharians' of antiquity (Strabon's Τόχαιοι; Skt. *Tukhāra*-). Among the figures in the spectacular Buddhist cave paintings of the region are some with red hair and green or blue eyes, and many have speculated that these were the Tocharians; more recently, the discovery of red-haired, 'Western'-looking mummies in the Taklamakan made headlines in the mid-1990s, but once again we cannot be sure which language they spoke. In any case, the speakers of Tocharian (more precisely, TB; see above) began to shift to Turkic in the later first millennium AD; the language was probably extinct by 1000.

Although they differ in numerous respects and certainly were not mutually intelligible, TA and TB were structurally similar, characterized by right-headed constituent phrases, a system of agglutinating

nominal case suffixes, and the central role of aspect and tense in verbal morphology. The two had doubtless been diverging for several centuries before the time of our documents, so that their latest reconstructible common ancestor, Proto-Tocharian (PT), must be dated to the last centuries BC.

The variety of Brāhmī script used to write Tocharian lacked symbols for distinctively Tocharian sounds such as labiovelar [k^w] and (in TB) the diphthongs [ew], [ow], [aw], [ay], and contained signs (*akṣaras*) for Sanskrit and Prakrit phonemes absent in the language (e.g., *ṽ*, *ḥ*, and the voiced, aspirated, and retroflex obstruents; the latter are used almost exclusively in Indo-Aryan borrowings). The most remarkable innovation of the Tocharian writing system is the creation of a series of *Fremdzeichen* ('foreign signs') to represent sequences of consonant + the vowel *ä*, probably a high central [i]; these exist for most, but not all consonants, and are used apparently interchangeably with the normal *akṣara* plus two subscript dots (hence the transcriptions *tä*, *ñä*, etc.). A second vowel (usually *u*) can be combined with a ligature (e.g., *kse* + *u*, transcribed *k_use*); such 'subscript' *u*'s may denote either a labiovelar or a reduced/syncopated vowel.

Recent research has elucidated most of the principal phonological developments from Proto-Indo-European (PIE) to PT and the two Tocharian languages. The PIE series of voiceless, voiced, and voiced aspirate stops have famously merged, except that *t, *d^h > PT *t remained distinct from *d > PT *t^s. PIE palatals and velars merged in PT, but labiovelars and sequences of palatal/velar + *w remained distinct. Palatalization before front vowels created new allophones that then became phonemic and gave rise to a number of morphologically conditioned alternations. The vowels underwent many changes, including loss of contrastive length. TB is in general the more phonologically conservative of the two languages, especially the western dialect; in contrast, TA has undergone sweeping changes, principally involving the vowel system.

The noun distinguishes two genders, masculine and feminine, plus a class of nouns of 'alternating' gender that take masculine agreement in the singular and feminine in the plural. Nouns and adjectives contrast for singular, plural, and dual. Nouns inflect for nine cases in each language, but only the three 'primary' cases, nominative, oblique, and genitive, are of PIE date; the remaining 'secondary' case suffixes are agglutinative, added to the oblique of singular and plural alike, and attached only to the last element of a noun phrase (e.g., TA *kuklas yukas oñkälmas-yo* 'with chariots, horses, and elephants'). Although their functions mostly coincide, few of the suffixes are

clearly cognate: cf. comitative TB *-mpa*, TA *-aśśäl*, ablative TB *-mem*, TA *-äs*. Most nonfeminine nouns have identical forms for nom. and obl. singular, derived from the PIE accusative, but masculine nouns denoting rational beings have secondarily created a distinct oblique in TB *-m*, TA *-(a)m*.

The numerals are of clear IE provenance: TB *wi*, TA m. *wu*, f. *we* '2,' TB m. *trey* ~ *trai*, f. *tarya*, TA m. *tre*, f. *tri* '3,' TB m. *štwer*, f. *štwärä*, TA *štwar* '4,' TB *piś*, TA *pāñ* '5,' TB *škas*, TA *šäk* '6,' TB *šukt*, TA *špät* '7,' TB *okt*, TA *okät* '8,' TB, TA *ñu* '9,' TB *šak*, TA *šäk* '10,' TB *ikäm*, TA *wiki* '20.' The prehistory of the personal and demonstrative pronouns contains a number of unsolved problems; noteworthy is the existence of separate masculine and feminine forms for 'I' in TA, m. *näš*, f. *ñuk* (vs. TB m./f. *ñäs*, *ñis*).

The verb exhibits numerous idiosyncratic developments alongside a wealth of interesting and archaic features and has played an increasingly prominent role in the ongoing debate over the reconstruction of the PIE verbal system. The inherited voice distinction of active and mediopassive is robustly preserved. An interesting, if often overemphasized, feature is the widespread suffixation of PT **-skē-* ~ **-ššə-* (> TB */-ske-/* ~ */-ššə-/* ~ */-s-/*; TA *-sa-* ~ *-š-*) to derive transitives to intransitive roots and causatives to many, but not all, transitive roots. Both languages have the same morphological categories of present and imperfect (= nonpast and past of the imperfective stem), subjunctive/future and optative (= nonpast and past of the perfective stem), imperative, and preterite; nonfinite forms include the infinitive, gerundives I and II (denoting respectively obligation and possibility), a verbal noun or 'abstract,' almost always built to gerundive II; and a present and preterite participle. Most inflectional categories and patterns of verbal stem derivation are of PIE date, including reflexes of nasal and stative presents and root and (pre-)sigmatic aorists. Approximately a dozen verbs are suppletive, second only to Old Irish among IE languages. Among numerous unsolved problems are the remarkable paucity of simple thematic presents, and the origin of the Tocharian subjunctive and its relation to the classical PIE subjunctive and perfect.

Nominal compounds are fairly common, as are complex derivatives like TB *raddhi-lak-ä-š(š-äly)-ñe-šše* 'of causing to see wonders.' As the Tocharian languages are left-branching, the verb is usually clause-final in prose documents but may be raised for various pragmatic effects; verse texts not surprisingly offer much variation.

Early on, many Indo-Europeanists were struck by the apparent connections between Tocharian and the western IE languages, particularly Celtic and

Germanic. Today, however, the emerging consensus holds that Tocharian is not closely related to any other branch of IE but, rather, was the second after Anatolian to diverge from the ancestral speech community.

Little is known of the earliest contacts between Tocharian and other languages. Several strata of Iranian loanwords may be distinguished: the oldest appear to date from the Old Iranian period, followed by a small set whose preforms strongly resemble Ossetic; most recent are loans from neighboring Eastern Middle Iranian languages, particularly Khotanese. A few old Indo-Aryan borrowings go back to the pre-PT period – note TB (prose) *pañäkte*, TA *ptāñkät* 'Buddha' and TA *pñi* 'puṇya', reflecting the change of **u > PT *ə* – but the huge number of loanwords from Sanskrit and Prakrit entered the language comparatively recently; some have been (partly) assimilated to Tocharian phonology, but many retain their original orthography and may have belonged only to high religious registers. Certain old loanwords in Chinese appear to come from Tocharian, for example, MidChin. **mjit* (or sim.) 'honey' ← PT **m^yətə* (cf. TB *mit*). In light of the Tocharians' ultimate linguistic shift to Turkic, it is interesting to note that a few important Turkic words may be of Tocharian origin: cf. TB *okso* 'ox', *kaum* (*kom*) 'day, sun' → Proto-Turkic **öküz*, **kün*.

Bibliography

- Adams D Q (1988). *Tocharian historical phonology and morphology*. American Oriental Series, vol. 71. New Haven, CT: American Oriental Society.
- Adams D Q (1999). *A dictionary of Tocharian B*. Leiden Studies in Indo-European 10. Amsterdam: Rodopi.
- Hackstein O (1995). *Untersuchungen zu den sigmatischen Präsensstambildungen des Tocharischen*. Historische Sprachforschung, Ergänzungsheft 38. Göttingen: Vandenhoeck und Ruprecht.
- Hilmarsson J (1996). *Materials for a Tocharian historical and etymological dictionary*. Tocharian and Indo-European Studies Supplementary Series, vol. 5. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Ji X, Winter W & Pinault G J (1998). *Fragments of the Tocharian A Maitreyasamiti-Nāṭaka of the Xinjiang Museum, China*. Trends in Linguistics, Studies and Monographs 113. Berlin: York: de Gruyter.
- Krause W (1952). *Westtocharische Grammatik. Band I: Das Verbum*. Heidelberg: Winter.
- Krause W & Thomas W (1960). *Tocharisches Elementarbuch. Band I: Grammatik*. Heidelberg: Winter.
- Lane G S (1963). 'On the interrelationship of the Tocharian dialects.' In Birnbaum H & Puhvel J (eds.) *Ancient Indo-European Dialects*. Berkeley: University of California Press. 213–233.
- Pinault G-J (1987). 'Épigraphie koutchéenne. I. Laissez-passer de caravanes. II. Graffites et inscriptions.' In

- Huashan C, Maillard M, Gaulier S & Pinault G (eds.) *Mission Paul Pelliot. Documents conservés au Musée Guimet et à la Bibliothèque Nationale. Documents archéologiques VIII: Sites divers de la région de Koutcha*. Paris: Collège de France, Instituts d'Asie, Centre de Recherche sur l'Asie Centrale et la Haute Asie. 59–196.
- Pinault G-J (1989). 'Introduction au tokharien.' *LALIES* 7, 3–224.
- Ringe D A Jr (1990). 'Evidence for the position of Tocharian in the Indo-European family?' *Die Sprache* 34, 59–123.
- Ringe D A (1996). *On the chronology of sound changes in Tocharian*. Vol. I: *From Proto-Indo-European to Proto-Tocharian*. American Oriental Series, vol. 80. New Haven, CT: American Oriental Society.
- Sieg E & Siegling W (1908). 'Tocharisch, die Sprache der Indoskythen. Vorläufige Bemerkungen über eine bisher unbekannte indogermanische Literatursprache.' *Sitzungsberichte der Berliner Akademie der Wissenschaften zu Berlin, Jahrgang 1908*, 915–934.
- Sieg E & Siegling W (eds.) (1949–53). *Tocharische Sprachreste: Sprache B. Heft 1: Die Udānālaṅkāra-Fragmente. Text, Übersetzung und Glossar* (1949). Heft 2: *Fragmente Nr. 71–633*. Auch dem Nachlaß herausgegeben von Werner Thomas (1953). Göttingen: Vandenhoeck und Ruprecht.
- Sieg E, Siegling W & Schulze W (1931). *Tocharische Grammatik*. Im Auftrage der Preußischen Akademie. Göttingen: Vandenhoeck und Ruprecht.
- Thomas W (1964). *Tocharisches Elementarbuch. Band II: Texte und Glossar*. Heidelberg: Winter.
- von Gabain A & Winter N (1958). *Türkische Turfantexte IX. Ein Hymnus an den Vater Mani auf 'Tocharisch' B mit alttürkischer Übersetzung*. Abhandlungen der Deutschen Akademie der Wissenschaften zu Berlin, Klasse für Sprachen, Literatur und Kunst, Jahrgang 1956, Nr. 2. Berlin: Akademie.
- van Windekens A J (1976–82). *Le tokharien confronté avec les autres langues indo-européennes. Volume I: Le phonétique et le vocabulaire* (1976). Volume II, Tome I: *La morphologie nominale* (1979). Volume II, Tome 2: *La morphologie verbale* (1982). Louvain: Centre International de Dialectologie Générale.

Toda

P Bhaskararao, Tokyo University of Foreign Languages, Tokyo, Japan

© 2006 Elsevier Ltd. All rights reserved.

Toda is the name of an ethnic group that resides on the Nilagiri mountains (= Nilgiris) in South India (with its major town of Udagamandalam = Ooty = Ootacamund located at 11.24°N and 76.44°E). The lofty Nilagiri mountains (with peaks rising above 2400 m) are the home of five culturally interrelated ethnic communities – Toda, Kota, Kurumba, Irula, and Badaga – all speaking different Dravidian languages. The Todas recognize the mutual relationship and the historical interdependency among these five communities. Though they are known as Todas to outsiders, they call themselves oṽ (meaning 'Toda person'), and their language, oṽ-poṣ (poṣ 'language'). The Toda language contains different words for each of the five Nilagiri communities. They are oṽ (Toda = tōḍa), *kuṽf* (Kōṭa), *kurb* (Kurumba), *erl* (Irula), and *maṽ* (Baḍaga). Though the Toda language does not have a word or a phrase to denote all these five communities together as one group, it has the word, *pōṛ*, which means 'a nonwhite low lander (who does not belong to any of the five Nilagiri communities.)' This term demarcates all the five communities together as a group from the rest of the people of India. A white person (who is expected not to be an Indian) is called *ars*.

Toda language is endangered, with just around 900 speakers, most of whom are bilingual in their mother tongue and in Tamil language, the most dominant language in the area. There is twofold division among the Todas, conventionally called 'moieties' in anthropology. The moieties are *toṛṭaṣ* and *tōwfiṽ*. Except for a few words and phrases, no significant dialectal differences are found between these two moieties.

Phonology

Among the languages of India, Toda possesses a unique and complex inventory of phonemes. It is the sound system of this language that makes it sound 'very foreign' to many non-Todas, making them to imagine wildly that it is because the Todas were originally some exotic people such as Greeks, Persians, and so forth their language contains so many 'foreign' sounds. All of these 'exotic' sounds are derivable historically from a proto-stage through an intricate set of rules.

In addition to the typical five vowels (and their long counterparts) available in most of the Dravidian languages, the Toda vowel inventory contains front rounded vowels, /ü, ö/, and a back unrounded vowel, /i/. In addition, each of the eight vowel positions has a short and long counterpart. This results in a total of 16 contrasting vowel phonemes viz., *i, ī, e, ē, ü, ǖ, ö, ȫ, i̇, i̇̄, u, ū, o, ō, a, ā*. The inventory of consonants presents the most complex system known for any Indian language, and some of the

consonantal contrasts found in this language are not encountered in any other language in the world. There are four voiceless sibilants contrasting at dental, alveolar, palatal, and retroflex places: /s, š, ʃ, ʂ/. There are three nonsibilant fricatives: /f, θ, x/. By some morphophonemic changes, the voiced counterparts of these fricatives also attain contrastive status. Voiceless and voiced plosives contrast at seven places of articulation. They are labial (p, b), dental (t, d), denti-alveolar (c, ɟ), alveolar (t̪, d̪), palato-alveolar (č, j), retroflex (ɖ, ɗ), and velar (k, g). There are three nasals: /m, n, ŋ/. This is the only known language that has a set of three contrastive trills and four contrastive laterals. The trills are dental /r/, alveolar /r̪/, and retroflex /ɽ/, and the laterals are voiceless alveolar /l̪/, voiced alveolar /l/, voiceless retroflex /ɭ/ and voiced retroflex /ɬ/ (Tables 1 and 2).

Sentences

Toda is an SOV language with syntax very similar to that of other Dravidian languages. The so-called ‘extra-subject predication’ (Emeneau, 1984: 51) is peculiar to this language; for example,

o n kify kōt-s-pini

I_{NOM} ear_{NOM} got.destroyed-PAST.SUFFIX-1ST
PERSON.SINGULAR

‘My ears were ruined’

In a parallel sentence in other Dravidian languages, a dative-subject and the verb in concord with the object are expected.

The reportative/quotative sentences are also different in this language. The subject of the embedded sentence is marked for accusative case; for example,

en-n “pod-či”īd, ka-k öštši/uncsi

I-ACC “will.come.3p” QUOT, crow said.3PERSON/
thought.3PERSON

‘The crow said/thought that I would come’

Pronouns and Nouns First- and second-person pronouns are differentiated for number and inclusiveness (inclusion or exclusion of the addressee). There is no differentiation on the basis of sex among the pronouns. The pronouns are: first person singular: *on*, exclusive plural *em*, inclusive plural, *om*; second-person singular *ni*, plural *nim*; third person *aθ* (and its plural *aθ-a'm* containing the plural suffix *-a'm*). Nouns are either simple or derived. Simple nouns are mono-morphemic, and derived nouns contain a suffix; for example, *kurb* ‘Kurumba community,’ *kurbč* ‘a Kurumba female.’

Nouns are inflected for plurality and case by means of suffixes. *-a'm* is the most common plural suffix (*ir* ‘buffalo,’ *ir-a'm* ‘buffalos’). Some of the case suffixes are accusative *-n*, dative *-k~g*, locative *-š*, ablative *-šn*, instrumental *-it*, causal *-id*, sociative *-wi*. Some of the postpositions are *pok* ‘at the time of,’ *taš* ‘above.’ Some case forms are also formed by adding post-positions.

Numerals and Modifiers

Formation of numerals follows the general Dravidian pattern. Numerals 1000, 100, and 1–10 are mono-morphemic. They are *wid* ‘1,’ *e-d* ‘2,’ *mu-d* ‘3,’ *no-ng* ‘4,’ *üz* ‘5,’ *o-r* ‘6,’ *öw* ‘7,’ *öt* ‘8,’ *winboθ* ‘9,’ *pot* ‘10,’ *no-r* ‘100,’ and *so-fer* ‘1000’. The formula for forming decades is 2, 3, and so on, followed by 10 (e.g., *mu-poθ* [3–10] ‘30’). The formula for series between decades (e.g., 31–39) is the numeral for decade followed by 1–9 (e.g., *mu-poθ-e-d* [3–10–2] ‘thirty-two’). Ordinals are derived from the above cardinals by the addition of the suffix *-o-θ* (e.g., *e-d-öθ* ‘second’).

As in the case of several other Dravidian languages, it is difficult to demarcate adjectives from nouns. A few descriptive adjectives are *kir/kin* ‘small,’ *per* ‘big,’ *poč* ‘green.’ Similarly, adverbs as a class consists of a very few members. Most of the forms that

Table 1 Vowels

| | <i>FU</i> | <i>FR</i> | <i>CR</i> | <i>CU</i> | <i>BU</i> | <i>BR</i> |
|------|-----------|-----------|-----------|-----------|-----------|-----------|
| High | i i* | ü ü* | | | ĩ ĩ* | u u* |
| Mid | e e* | ö ö* | | | | o o* |
| Low | | | | a a* | | |

Table 2 Consonants

| | <i>Labial</i> | | <i>Dental</i> | | <i>Denti-Alveolar</i> | | <i>Alveolar</i> | | <i>Palato-Alveolar</i> | | <i>Retroflex</i> | | <i>Velar</i> |
|---------------------|---------------|-----|---------------|-----|-----------------------|-----|-----------------|-----|------------------------|-----|------------------|-----|--------------|
| <i>Stops</i> | p | b | t | d | c | ɟ | t̪ | d̪ | č | j | ɖ | ɗ | k g |
| <i>Nasals</i> | | m | | | | | | n | | | | ɳ | |
| <i>Fricatives</i> | f | (v) | θ | (ð) | s | (z) | š | (ž) | š | (ž) | ʃ | (ʒ) | x (y) |
| <i>Trills</i> | | | | r | | | | r̪ | | | | r̪ | |
| <i>Approximants</i> | | | | | | | | | | y | | | w |
| <i>Laterals</i> | | | | | | | l̪ | l | | | ɭ | ɬ | |

function as adverbs are inflected nouns or forms derived from verbs. A few clear cases of adverbs are *maxar* 'earliest,' *p̄in* 'later, afterwards.'

Oblique Forms

Some nouns, numerals, and so on are converted into oblique forms when some case suffixes are added to them. The most common oblique suffix is *-t*; for example, *me'n* 'tree': *me'n-t-k* 'to the tree' [*-k* is the dative suffix]; *e'd* 'two': *n̄im e'Ł-k* 'to you both.'

Verbs

The structure of Toda verb is quite complex compared to the rest of Dravidian languages. A verb is either simple (containing only a verbal root) or derived (root+derivative suffix). The most common derivative suffix is a transitive/causative suffix (e.g., *n̄il-* 'to stand,' *n̄il-c-* 'make to stand'). In a majority of cases, the transitive/causative suffix fuses with the ending of the root, as in *o'r-* 'to become dry,' *o't-* 'to dry something.' Mediative forms can be derived from a simple or derived base by addition of the suffix *-et̄~ety-*. The mediative form denotes a type of indirect or noncontactual causation; for example, *miy~mis-* 'to graze (intr),' *m̄i'c-~m̄i'č-* 'to graze (tr)' occurring in: *ir-a-n a tiṭ-ār m̄i'c* '(Go and) graze (tr.) the buffalos over that hill!'; *mors-fy ir-k willy mad kw̄irt miy-e't* 'Give some good medicine to the buffalos with foot-and-mouth disease and make them graze!'

Almost all of the Toda verbal bases have two morphophonemic alternants, conventionally called Stem 1 (S1) and Stem 2 (S2). S2 enters into a majority of inflections. Historically it corresponds to the past-tense stem in some of the South Dravidian languages (such as Tamil). S1 is the etymologically underlying form, and the corresponding S2 is derivable by a set of rules from it. For instance, 'to stand' -S1: *n̄il-* > S2: *n̄id-*. In addition, there is a third variety of stem called the 'Desiderative' stem that is peculiar to Toda among the Dravidian languages. The desiderative stem takes some inflectional suffixes.

A verb base alone functions as the singular imperative form; for example, *part* 'You(sg.) pray!.' All other full verbal forms contain a verb base followed by an optional inflection layer for tense/mode. A further optional inflectional layer of Pronominal suffixes (PN) that reflects the pronominal class of the subject terminates the verb. The third-person PN suffix is selected if the subject is any noun (singular or plural) or if it is a third-person pronoun (singular or plural). When the subject is a personal pronoun, the corresponding PN suffix is chosen. The PN layer is very complex in Toda compared with other Dravidian languages. It contains several sets of PN suffixes

that are added to different inflections. In the case of some inflections such as Non-Past, just the verb base followed by the appropriate PN suffix without an intervening tense suffix functions as the full verb. Another complexity of the PN suffix layer is the occurrence of two morphophonemic alternants of the same suffix across two sets that are conventionally called Paradigm I and Paradigm II. Paradigm I occurs before a terminating declarative suffix *-i*, and Paradigm II elsewhere.

Examples of a few tenses/modes are listed here (the verb base *t̄in-~t̄id-* 'to eat' is followed by one or more of the following suffixes: T=tense/mode suffix, Pn=Person suffix, D=declarative suffix: Past: *t̄id-s-s-i* [T-Pn-D] 'He/she/it/they ate'; Non-past: *t̄id-č-i* [Pn-D] 'He/she/it/they will eat'; Negative: *t̄in-in-i* [Pn-D] 'I did/do/will not eat'; Voluntative: *t̄in-g-y* [T-Pn] 'You(sg) may eat'; Tenseless: *t̄id-en* [T-Pn] 'I eat'; Plural Imperative: *t̄in-š* [Pn] 'You (pl.) eat!'; Conditional : *t̄id-u-fir* [Pn-M] 'If he/she/it eats'; Contemporaneity : *t̄id-u-k* [Pn-M] 'while we (incl) were eating'; *t̄id-pok* [T-Pn] 'when (somebody) ate'; Purposive: *t̄id-p̄ik* [T-P] 'for the sake of eating.'

A number of verb bases also function as auxiliary verbs (Ax) in producing various modal forms. Some examples are: *kuty-īs-s-pini* [*kuty-* 'to embroider'-Ax-T-Pn] 'I knew how to embroider'; *pod-k̄iṣ-īyi* [*pod-* 'to come'-Ax-Pn] 'He could not come'; *noby-p̄it-s-pini* [*noby-* 'to believe'-Ax-T-Pn] 'I believed (him) wrongly,' *t̄id-kur̄y-s-py* [*t̄id-* 'to eat'-Ax-T-Pn] 'You(sg) have completed eating'; *afoṭ-kw̄id-iti* [*afoṭ-* 'to talk'-Ax -Pn] 'Don't talk to yourself,' *k̄is-s-pod-s-si* [*k̄is-* 'to do' -Ax-T-Pn] 'He went on doing,' *kōṭ-s-pi-č-či* [*kōṭ* 'to be spoiled' -T-Ax-Pn] 'It will get spoiled.'

A set of suffixes derives verbal forms that are used as modifiers; for example, *t̄id-fy o'Ł* 'man who ate'; *t̄id-θ o'Ł* 'man who ate (more definite)'; *t̄id-t o'Ł* 'man who eats'; *t̄id-p o'Ł* 'man who can eat'; *t̄in-o-fy o'Ł* 'man did/does not eat.' The suffix *-t* also derives a verbal noun as in: *nōw k̄is-t* 'making a song' [*k̄iy-~k̄iṣ-* 'to make, do'].

Vocabulary

Before the Todas came into contact with words from the languages from the plains such as Tamil, Hindi, and English, their borrowed vocabulary must have consisted of a few words from the language of their neighbors such as Badagas. In contemporary Toda language, there are a few words from Badaga (and some Indo-Aryan words borrowed through Badaga), Tamil, Hindi, and English. Because of the elaborate ritual structure, Toda language has developed an intricate system of naming of persons, water buffalos, and places. Like other languages, it has

some culture-specific vocabulary—for instance, because of the importance of the water buffalo in the spiritual as well as mundane planes of their lives, we find words with very specific meanings such as: *malf* ‘buffalo gives a side glance before attacking.’ The traditional songs (some of them are presently in their twilight stage) contain special words as well as word-formation processes. The most important component of a Toda song is a paired unit called *koŋ* (called song-units by earlier authors). An example of a *koŋ* is: *poŋ-terθ-poŋm # twi-terθ-ir* [box-having.opened-money # buffalo.pen-having.opened-buffalos] (just open the box and give money or open the buffalo-pen and give buffalos to anybody who asks>) ‘to behave very generously.’

Bibliography

- Emeneau M B (1958). ‘Oral poets of South India – the Todas.’ In Hymes D (ed.) *Language in culture and society*. New York: Harper and Row. 330–341.
- Emeneau M B (1965). ‘Toda dream songs.’ *Journal of American Oriental Society* 85, 39–44.
- Emeneau M B (1965). ‘Toda verbal art and Sanskritization.’ *Journal of the Oriental Institute, Baroda* 14, 273–279.
- Emeneau M B (1966). ‘Style and meaning in an oral literature.’ *Language* 42, 323–345.
- Emeneau M B (1971). *Toda songs*. London: Oxford University Press.
- Emeneau M B (1974). *Ritual structure and language structure of the Todas*. Philadelphia: Transactions of the American Philosophical Society. New Series – 64:6.
- Emeneau M B (1979). ‘Linguistic archaisms in Toda songs.’ *South Asian Language Analysis (SALA)* 1, 41–45.
- Emeneau M B (1984). *Toda grammar and texts*. Philadelphia: American Philosophical Society.
- Nara T & Bhaskararao P (2001). *Toda vocabulary: a preliminary list*. Osaka: ELPR Publications Series A3–002.
- Nara T & Bhaskararao P (2002). *Toda texts*. Osaka: ELPR Publications Series A3–005.
- Nara T & Bhaskararao P (2003). *Songs of the Toda*. Osaka: ELPR Publications Series A3–001.
- Shalev M, Ladefoged P & Bhaskararao P (1994). ‘Phonetics of Toda.’ *PILC Journal of Dravidic Studies* 4, 19–56.
- Spajić S, Ladefoged P & Bhaskararao P (1996). ‘The trills of Toda.’ *Journal of the International Phonetic Association* 26, 1–21.

Tohono O'odham

M Miyashita, University of Montana, Missoula, MT, USA

© 2006 Elsevier Ltd. All rights reserved.

Tohono O'odham ‘Desert People,’ formerly known as Papago, belongs to the Tepiman (or Pimic) branch of the Uto-Aztecan language family, and is closely related to the Akimel O'odham (or Pima ‘River People’). O'odham is spoken in Sonora, Mexico and Southwestern Arizona (Tohono O'odham, San Xavier, Ak Chin, Gila River, and Salt River). The estimated number of speakers is between 14 000 and 15 000 (Zepeda p.c. in Mithun, 1999).

In the early 1900s, Juan Dolores, a native O'odham speaker, documented the language with linguist J. Alden Mason (Mathiot, 1991). Dolores published collections of O'odham verbs (Dolores, 1913), noun stems (Dolores, 1923), and nicknames (Dolores, 1936). Mason and Dolores compiled their work into an O'odham grammar. Ken Hale (1959) wrote his dissertation, ‘A Papago Grammar,’ based on fieldwork with O'odham speakers, and Dean Saxton (1963) published an article on the O'odham phonemic system. Madeline Mathiot (1973) compiled an extensive dictionary with the grammatical usage. Saxton *et al.* (1983) published an English-O'odham/Pima and

Pima/O'odham-English dictionary. Zepeda (1984), a native speaker, wrote a pedagogical grammar of Tohono O'odham. Some books portray O'odham songs (Bahr *et al.*, 1997; Underhill, 1993). A number of recent articles in O'odham discuss the language from theoretical points of view (Hale, 1983; Hill and Zepeda, 1992; Fitzgerald, 1997, 1998, 2000, 2002; Miyashita, 2002; Truckenbrodt, 1999).

Many loanwords are from Spanish, and some are from other indigenous languages (Miller, 1990; Hill, 1998). Five major dialects are recognized: Totoguañ, Koló:di, Gigimai, Hú:hu'ula, Ko:adk, and Huhuwos (S'óobemakame) (Saxton, 1963; Saxton *et al.*, 1983). There are generational and gender variations. Sentential conjunction such as *kuñ* ‘... and I ...’ and *kup* ‘... and you ...’ may be shortened to *n* and *p* (Zepeda, 1983). The former is more formal than the latter. Women use inhalation, or pulmonic ingressive airstream, in discourse for intimate interactional purposes (Hill and Zepeda, 1999).

The Alvarez-Hale writing system (Alvarez and Hale, 1970) is the official orthography of the Tohono O'odham Nation (Zepeda, 1983). Dictionaries by Mathiot (1973) and Saxton *et al.* (1983) use their own systems and are different from the Alvarez-Hale system (Zepeda, 1983; Miyashita and Moll, 1999).

Zepeda (1983) describes that O'odham exhibits 19 consonants: *b*, *c* (= *tʃ*), *d*, *ɖ* (= *d*), *g*, *h*, *j* (= *dʒ*), *k*, *l* (= *ʎ*), *m*, *n*, *ɲ*, *ŋ*, *p*, *s*, *ʂ* (= *ʃ*), *t*, *w*, *y* (= *j*), and asymmetric five vowels *i*, *e* (= *ɪ*), *u*, *o*, *a*. Although minimal pairs are rarely found, long and short vowels are phonemic (Hale, 1959). For example, *hik* 'navel' vs. *hi:k* 'cut,' *ta:tk* 'feel' vs. *tatk* 'the root of a plant.' There are extra-short (or aspirated, voiceless) vowels marked with a breve [i] in the orthography. An additional example, *toki* 'cotton' vs. *go:kɪ* 'footprint.' Allophonic variations appear in native words. Phonemes *t d n ɖ*, and *ʂ* appear as *c j ɲ l*, and *s* before *li*.

The stress falls on an initial syllable, e.g., *músiɡo* 'musician.' Prefixes do not bear a stress, e.g., *ha-wápkon* 'them-wash.' Secondary stress appears in polymorphemic words (Fitzgerald, 1997). Some loanwords are noninitially stressed, e.g., *paló:ma* 'dove < Spanish *paloma*.'

Nouns and verbs undergo partial reduplication for plural and/or distributive indications, e.g., *gogs* → *gogogs* 'dog(s),' *him* 'walking sing.' → *hibim* 'walking pl.', etc. Some words do not reduplicate, e.g., *cicwi* 'playing.' (Zepeda, 1984; Hill and Zepeda, 1998). Truncation forms a perfective verb by dropping the last consonant of an imperfective verb, e.g., *o'ohan* 'writing' → *o'oha* 'wrote.' This does not apply to a vowel-final word, e.g., *cicwi* 'playing' → *cicwi* 'played,' *si'i* 'sucking' → *si:* 'sucked.'

Person/number of possessives is indicated by a prefix, except for third person singular, which is a suffix (Zepeda, 1984). Possessed nouns are classified into alienable and inalienable categories. Alienable nouns are analyzed as either they having been previously unowned or as being related to sequential human ownership (Bahr, 1986). As shown in (1) and (2), an alienable noun must have the suffix *-ga*, and an inalienable does not.

- (1) *ɲ-je'e*
1.sing.POSS-mother
'my mother'
- (2) *gogs-ga-j*
dog-alienable-3.sing.POSS
'his or her dog'

O'odham is a nonconfigurational language. All six orders (SOV, SVO, OSV, OVS, VSO, and VOS) are possible for a transitive sentence without meaning alteration (Zepeda, 1984; Miyashita *et al.*, 2003). Pragmatic status may correlate the word order (Payne, 1987). Any nonpronominal noun must follow a particle called a *g*-determiner, except when it is sentence-initial. A sentence must have an auxiliary (AUX), which must be in the second position of the sentence as only one restriction regarding the configuration (Zepeda, 1984).

- (3) Wakial 'o ceposid g haiwan. (SVO)
cowboy. AUX. branding DET cow.
sing 3.sing sing
'The cowboy is/was branding the cow.'

- (4) Ceposid 'o g wakial g haiwan. (VSO)
branding AUX. DET cowboy. DET cow.
3.sing sing sing
'The cowboy is/was branding the cow.'

AUX indicates the subject's person and number, and a verb prefix indicates that of the object. However, as shown in sentences (3) and (4), any third person singular argument has no overt indication regarding the grammatical relation. Although there is no overt case marking, O'odham may be an ergative language because a reduplicated intransitive verb agrees with the plural subject, while a reduplicated transitive verb agrees with the plural object in a sentence (Zepeda, 1983; Miyashita, 2002).

Verbal aspects are distinguished between completed and continued actions (Dolores, 1913; Zepeda, 1983). Tense is divided into future and nonfuture. Imperfective present and past are not grammatically distinguished. Future is marked by a particle *o* with perfective AUX. Future imperfective sentences are formed with the perfective AUX, the imperfective verb, and usually the suffix *-d/-ad* on the verb.

O'odham kin terms show the equilateral system (Saxton *et al.*, 1983). Siblings and cousins are the same, *ʂe:pij*. Parents' siblings have eight distinct terms depending on the gender, age, and lineage. Grandparents have four distinct terms. Great-grandparents have one term, *wi:kol*. Great-great-grandparents have the same term as siblings, *ʂe:pij*. The term for 'child' is distinct depending on the gender of the parent rather than of the child.

Bibliography

- Alvarez A & Hale K (1970). 'Toward a manual of Papago grammar: some phonological terms.' *International Journal of American Linguistics* 36(2), 83–97.
- Bahr D M (1986). 'Pima-Papago-ga, "Alienability."' *International Journal of American Linguistics* 52(2), 161–171.
- Bahr D M, Paul L & Joseph V (1997). *Ants and orioles: showing the art of Pima poetry*. Salt Lake City: University of Utah Press.
- Dolores J (1913). 'Papago Verb Stems.' *American Archaeology and Ethnology* 10, 241–263.
- Dolores J (1923). 'Papago Noun Stems.' *American Archaeology and Ethnology* 20, 19–31.
- Dolores J (1936). 'Papago Nicknames.' In Mason J A (ed.) *Essays in anthropology in honor of Alfred Kroeber*. Berkeley: University of California Press. 45–47.
- Hale K (1959). 'A Papago grammar.' Ph.D diss., Indiana University.

- Hale K (1983). 'Papago (k)c.' *International Journal of American Linguistics* 49(3), 299–327.
- Hill J (1998). [Journal issue dated 1993, submitted 1997] 'Spanish in the indigenous languages of Mesoamerica and the southwest: beyond stage theory to the dynamics of incorporation and resistance.' *Southwest Journal of Linguistics* 12(1–2), 87–108.
- Hill J & Zepeda O (1992). 'Derived Words in Tohono O'odham.' *International Journal of American Linguistics* 58, 355–404.
- Hill J & Zepeda O (1998). 'Tohono O'odham (Papago) Plurals.' *Anthropological Linguistics* 40(1), 1–42.
- Hill J & Zepeda O (1999). 'Language, gender and biology: pulmonic ingressive airstream in women's speech in Tohono O'odham.' *Southwest Journal of Linguistics* 18, 15–40.
- Fitzgerald C M (1997). 'O'odham Rhythm.' Ph.D. diss., University of Arizona.
- Fitzgerald C M (1998). 'The meter of Tohono O'odham songs.' *International Journal of American Linguistics* 64, 1–36.
- Fitzgerald C M (2000). 'Vowel Hiatus and faithfulness in Tohono O'odham reduplication.' *Linguistic Inquiry* 31, 713–722.
- Fitzgerald C M (2002). 'Tohono O'odham stress in a single ranking.' *Phonology* 19, 253–271.
- Mathiot M (1973). *A dictionary of Papago Usage* (vols 1 & 2). Indiana: Indiana University Publications.
- Mathiot M (1991). 'The reminiscence of Huan Dolores, an Early O'odham linguist.' *Anthropological Linguistics* 33, 233–315.
- Miller W (1990). 'Early Spanish and Aztec loan words in the indigenous languages of Northwest Mexico.' In Cuaón B G & Levy P (eds.) *Homenaje a Jorge A Suárez: Lingüística Indoeuropea e Hispanica*. Mexico, D. F.: El Colegio de México. 351–366.
- Mithun M (1999). *The Languages of North America*. Cambridge: Cambridge University Press.
- Miyashita M (2002). 'Tohono O'odham Syllable Weight: Descriptive, Theoretical and Applied Aspects.' Ph.D. Dissertation. University of Arizona.
- Miyashita M & Moll L (1999). 'Enhancing Language Material Availability Through Computer Technology.' In Reyhner J, Cantoni G & St. Clair R N (eds.) *Revitalizing Indigenous Languages*. Flagstaff: Northern Arizona University. 113–116.
- Miyashita M, Demers R & Ortiz D (2003). 'Grammatical Relations in Tohono O'odham: An Instrumental Perspective.' In Karimi S (ed.) *Word Order and Scrambling*. Oxford: Blackwell. 44–66.
- Payne D L (1987). 'Information Structuring in Papago Narrative Discourse.' *Language* 63, 783–804.
- Saxton D (1963). 'Papago Phonemes.' *International Journal of American Linguistics* 29, 29–35.
- Saxton D, Saxton L & Enos S (1983). *Dictionary: Papago/Pima-English, English-Papago/Pima*. Tucson: The University of Arizona Press.
- Truckenbrodt H (1999). 'On the relation between syntactic phrases and phonological phrases.' *Linguistic Inquiry* 30, 219–256.
- Underhill R M (1993). *Singing for power*. Tucson: The University of Arizona Press.
- Zepeda, Ofelia (1983). *A Papago grammar*. Tucson: The University of Arizona Press.
- Zepeda, Ofelia (1984). 'Topics in Papago morphology.' Ph.D. diss., University of Arizona.

Tok Pisin

S Romaine, Oxford University, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Tok Pisin (from English *talk pidgin*) is an English-lexicon pidgin spoken by approximately three-quarters of Papua Guinea's approximately 5 million inhabitants. It is not only the lingua franca of the entire country, with its 800 some indigenous languages, but it is also the language spoken by the most people in the South Pacific today. It is closely related to and mutually intelligible with Pijin in the Solomon Islands and Bislama in Vanuatu. All three varieties of Melanesian Pidgin owe their origins to the Queensland sugarcane plantations to which as many as 100 000 workers from these three countries were recruited during the 19th century. Men with mutually unintelligible village languages found themselves living and working together, as well as needing to

communicate with their English-speaking plantation managers. A form of pidgin English served this purpose. When the labor trade ended in 1905, most of the workers went back to their countries of origin, taking with them knowledge of this Queensland Plantation Pidgin. In these highly multilingual countries, pidgin served the useful internal function of communicating across ethnolinguistic boundaries. Social conditions were thus conducive not just for the retention and spread of the pidgin but also for its stabilization and subsequent creolization.

Today, Tok Pisin is used across Papua New Guinea's social spectrum, known by villagers and government ministers alike. It is the most frequently used language in the House of Assembly, the country's main legislative body, and the constitution recognizes Tok Pisin as one of the national languages of Papua New Guinea. Tok Pisin has become the main language of the migrant proletariat and the first language of the

younger generation of town-born children, where it has creolized (i.e., become a creole). Tok Pisin is also one of the few pidgin and creole languages to have undergone considerable standardization because missionaries realized its potential early on as a valuable lingua franca for proselytizing among a linguistically diverse population and began using it for teaching. Most printed material is still religious; the Bible has been translated into Tok Pisin. However, the language is used to some extent in radio and television broadcasting, especially in interviews and news reports. The weekly Tok Pisin newspaper, *Wantok*, has a readership of over 30 000. Until recently, English was the only official language of education in Papua New Guinea despite the fact that few children enter school knowing it. However, education reforms have allowed communities to choose the language to be used in the first 3 years of elementary education, and many have chosen Tok Pisin.

The lexicon of Tok Pisin is mainly English (79%); Tolai (Kuanua), an indigenous language, has contributed 11%, other indigenous languages 6%, German 3%, and Malay 1%; there is also a handful of words from other European languages such as Portuguese/Spanish (e.g., *save* from Portuguese/Spanish *sabir/saber* 'to know/knowledge'). English borrowings provide the most important source of new vocabulary. Even frequently used words such as *kiau* 'egg' from Tolai are increasingly being replaced by or used alongside the English *egg*. Likewise, some of the German vocabulary in the language (e.g., *beten* 'pray'), dating from the period of German colonial rule (1884–1914) of part of the country, is giving way to English. The phonology of individual speakers of Tok Pisin varies from a core system that is shared by all speakers of the language and is similar to that of the indigenous substrate languages to a highly anglicized phonology that makes the most of English consonant and vowel distinctions. Tok Pisin has little morphology, although it has acquired some derivational and inflectional morphology in the course of its expansion. The suffix *-pela* is used to form the plural of the first- and second-person pronouns (e.g., *yu-pela* 'you plural'); it also marks a subset of monosyllabic attributive adjectives, demonstratives, and cardinal numerals (e.g., *dis-pela tu-pela pis* 'these two fish(es)'). Transitive verbs are marked with the suffix *-im*.

| | | | | |
|--------|------|-------------|----------|---------------|
| em | i | lus-im | dis-pela | ples |
| 3.SING | PRED | leave-TRANS | this-DEM | place/village |

'he/she/it left this place/village'

(Here, PRED stands for predicate marker.) Some grammatical distinctions reflect the influence of the substrate indigenous languages; in the personal pronoun

system, different pronouns are used for inclusive (i.e., speaker + hearer) and exclusive (i.e., speaker + other(s), not including hearer). Compare *yumi* 'we (inclusive)' with *mipela* 'we (exclusive)'.

Full lexemes are often used to express grammatical categories such as case, number, gender, tense, mood, and aspect, which in other languages are expressed by inflectional morphology; these distinctions are, however, not always obligatory. This is especially true for tense, mood, and aspect. The normal way of indicating past time is through use of the unmarked verb form, but *bin* (from English *been*) may be used.

| | | | | | |
|------|------|------|-------------|--------|------|
| ol | i | bin | adopt-im | liklik | meri |
| 3.PL | PRED | PAST | adopt-TRANS | little | girl |

'they adopted a/the little girl'

Sometimes *bin* is used in conjunction with *pinis* (from English *finish*) to mark completed actions in the past as a kind of perfective marker.

| | | | | | | |
|------|--------|---------|------|------|-----------|-------|
| tim | bilong | Mormads | i | bin | win-im | pinis |
| team | of | Mormads | PRED | PAST | win-TRANS | PERF |

'the Mormads team has won (the grand netball final)'

The meanings of immediate and remote future, prediction, intention, and irrealis may be expressed by clause-initial or preverbal *bai*, which has almost entirely replaced the earlier form *baimbai* (from English *by and by*). Pronominal subjects tend to take clause-initial *bai*, and noun phrases tend to take preverbal *bai*. Preverbal position is fast becoming the preferred order, although there are regional and stylistic differences.

| | | |
|--------|-----|----|
| mi | bai | go |
| 1.SING | FUT | go |

'I will go'

Tok Pisin has SVO word order; there is no copula and negation is preverbal (*mi no save* 'I don't know'). There is no inversion for questions.

| | | | |
|---------|------|---------|----|
| yu-pela | gat | brus | a? |
| 2-PL | have | tobacco | Q |

'do you (plural) have tobacco?'

Bibliography

- Mühlhäusler P, Dutton T E & Romaine S (eds.) (2003). *Varieties of English around the world: Tok Pisin texts: from the beginning to the present*. Amsterdam: John Benjamins.
- Romaine S (1992). *Language, education and development: rural and urban Tok Pisin in Papua New Guinea*. Oxford: Oxford University Press.
- Wurm S A & Mühlhäusler P (eds.) (1985). *Pacific linguistics C-70: Handbook of Tok Pisin*. Canberra, Australia: Australian National University.

Torricelli Languages

M Donohue, National University of Singapore, Singapore

© 2006 Elsevier Ltd. All rights reserved.

The approximately 50 languages of the Torricelli are spoken in north Papua New Guinea. The family extends from the eastern Bewani mountains in Sandaun Province; through the Torricelli ranges to Maprik, where Ndu speaking villages reach through to the north coast; and continuing east of Wewak in Sepik Province in the Marienberg ranges and ground south of the Murik lakes, with a final outpost at Bogia in Madang province. The languages are remarkable for non-Austronesian languages in New Guinea for having a basic SVO word order, whereas the norm is SOV. They have been grouped into seven subgroups, whose internal constituency appears to be valid, although the seven-way division still awaits proof. The membership of the family as a whole appears to be accurate.

There are typically no phonetically unusual segments in Torricelli languages, and, although stress is frequently contrastive, reports of tonal differences are rare. The languages near Nuku share with the adjacent Ndu languages the presence of creaky or glottalized vowels, ranging from just one (/a/) to contrasts present on the whole vowel inventory. The vowel inventories tend to be large, with seven or eight vowels being not uncommon in the western languages (a typical inventory is /i e ɛ a ɔ o u ʉ/) and five or six vowels being more common in the east. The loss of velar segments in some western languages has led to the unusual case of languages without velar contrasts at all. Voicing contrasts are usually associated with prenasalization.

There is significant diversity within the family, and the Torricelli languages are also significantly different from most other languages of New Guinea. Although they all show SVO order, typically with prefixal agreement for the subject and suffixal agreement for the object and lacking case marking on (core) nominals – all features that are unusual in New Guinea – other details of their morphological and syntactic structure show considerable diversity. In the eastern languages, such as Monumbo and Arapesh (Bukiyip, also known as Muhiang), multiple class systems with extensive concord are found, whereas in the west only remnant traces of noun classification can be found in the synchronically irregular plural endings of One and Olo.

For example, in Bukiyip ‘stone’ is *utom* (SING) *utabal* (PL), showing the *-m* and *-bal* suffixes typical of class 5 nouns (compare this with a class 2 noun, such as ‘village’ *wa-bél* SING, *wa-lúb* PL). Adjectives show

similar suffixes, agreeing in class and number with their noun, and verbs have cognate prefixes:

yopi-**mi** uto-**m** m-a-pwe agnú
‘(the) good stone is there’

yopi-**bili** wa-**bél** bl-a-pwe agnú
‘(the) good village is there’

In the western Torricelli language One, ‘stone’ is *toma* (SING) *tomu* (PL), showing an *-a* versus *-u* pattern, just as in ‘flower’ *sula* (SING), *sulu* (PL), indicating that, although it is a minority pattern, the alternations in ‘stone’ are regular. The word for ‘village’ *wapli* can be singular or plural, with the common *-li* plural suffix, but *wap* is only singular (this form is commonly found in compounds, such as *wap oi* ‘village grounds, area’). Concord on other words is not as strong, however:

upo toma w-ae nu
‘the good stone is there’

This sentence shows no agreement on *upo* ‘good,’ and only the general second/third-person singular *w-* on the verb ‘sit, be at.’ The same forms as are found in:

upo wapli w-ae nu
‘the good village is there’

A few adjectives do show alternations:

plola toma w-ae nu
‘the short stone is there’

plolu tomu n-ai n-e nu
‘the short stones are there’

with variation for number (the verb ‘sit’ has irregular singular and plural forms). Different noun classes, however, do not show different agreement patterns. Using the same inflecting adjective, *plola*, with a different noun shows the same inflectional pattern:

plola wap w-ae nu
‘the short village is there’

plolu wapli n-ai n-e nu
‘the short villages are there’

There are also no differences in verbal morphology. Another striking aspect of the NP in One involves the lack of a fixed word order: Gen N as well as N Gen, Dem N as well as N Dem, and Adj N as well as N Adj are found, with only relative clauses being restricted to postnominal position.

Like most languages of New Guinea, there is no evidence of a voice system operating in any of the Torricelli languages, but applicatives are almost universal in the Torricelli languages, being found in at least fossilized form even on the more isolating

members of the family. In some languages the applicative and the verb 'give' show close similarities (One: *-ne* APPL and *an(e)* 'give'), whereas in other languages the two morphemes bear no obvious resemblance to each other (Olo: *-f(i)* APPL, *wa* 'give'; Arapesh *-ma* APPL, *se* 'give'). There does not seem to be a single historical source for the various applicatives attested in different branches of the family. An applicative is often required lexically by low-transitive verbs. One has *y-upa-ne* 'follow,' with a lexicalized applicative, for instance.

Serial verbs are a regular feature of Torricelli languages, although clause chaining is not. One, the westernmost Torricelli language has an unusual syntactic parameter setting whereby word order within the NP is free but the position of NPs and PPs within the clause is rigidly fixed, implying that there is configurationality at the clause level but not at the phrase level.

Over the years, there have been various suggestions concerning the history of the Torricelli languages. Authors have suggested a relationship with the Asli languages of Malaysia and with the East Bird's Head languages of western New Guinea. None of these claims has yet stood up to any serious investigation. The SVO order of the Torricelli languages, unusual in New Guinea, has been attributed to Austronesian contact (as has also been proposed for the similarly SVO languages of the Bird's Head), but it could just as easily be innate. The Torricelli languages are, indeed, not highlands languages, and there is no reason to suppose that SVO is not the original Torricelli order.

Bibliography

- Conrad R J (1978a). 'Some Muhiang grammatical notes.' In *Workpapers in Papua New Guinea languages 25: Miscellaneous papers on Dobu and Arapesh*. Ukarumpa, Papua New Guinea: SIL. 89–130.
- Conrad R J (1978b). 'A survey of the Arapesh language family of Papua New Guinea.' In *Workpapers in Papua New Guinea languages 25: Miscellaneous papers on Dobu and Arapesh*. Ukarumpa, Papua New Guinea: SIL. 57–77.
- Conrad R & Wogiga K (1991). *An outline of Bukiyip grammar*. Canberra, Australia: Pacific Linguistics.
- Crowther M (2001). All the One language(s): linguistic and ethnographic definitions of language. Ph.D. diss., University of Sydney.
- Donohue M (2000). 'One phrase structure.' In Allan K & Henderson J (eds.) *Proceedings of ALS2k, the 2000 Conference of the Australian Linguistic Society*. University of Melbourne. Available at: <http://www.arts.monash.edu.au>.
- Foley W A (1986). *The Papuan languages of New Guinea*. Cambridge, UK: Cambridge University Press.
- Fortune R F (1942/1977). *Publications of the American Ethnological Society 19: Arapesh*. New York: AMS.
- Gerstner A (1963). *Micro-Bibliotheca Anthropos 37: Grammatik der Alübensprache*. Sankt Augustin, Germany: Anthropos Institute.
- Klafl J & Vormann F (1905). 'Die Sprachen des Berlinhafen-Bezirks in Deutsch-Neuguinea.' *Mitteilungen des Seminars für Orientalische Sprachen* 8, 1–138.
- Laycock D C (1975). 'The Torricelli phylum.' In Wurm S A (ed.) *New Guinea area languages and language study 1: Papuan languages and the New Guinea linguistic scene*. Canberra, Australia: Pacific Linguistics. 767–780.
- Manning M & Saggars N (1977). 'A tentative phonemic analysis of Ningil.' *Workpapers in Papua New Guinea Languages* 19, 49–71.
- Martens M & Tuominen S (1977). 'A tentative phonemic statement in Yil in West Sepik Province.' *Workpapers in Papua New Guinea Languages* 19, 29–48.
- McGregor D E & McGregor A R F (1982). *Olo language materials*. Canberra, Australia: Pacific Linguistics.
- Sanders A G & Sanders J (1980a). 'Dialect survey of the Kamasau language.' In Boxwell M, Goddard J, Ross M, Sanders A G, Sanders J & Davies J (eds.) *Papers in New Guinea Linguistics*, vol. 20. Canberra, Australia: Pacific Linguistics. 137–190.
- Sanders A G & Sanders J (1980b). 'Phonology of the Kamasau language.' In Boxwell M, Goddard J, Ross M, Sanders A G, Sanders J & Davies J (eds.) *Papers in New Guinea Linguistics* 20. Canberra, Australia: Pacific Linguistics. 111–136.
- Schmidt W & Vormann F (1900). 'Ein Beitrag zur Kenntnis der Valman-Sprache.' *Zeitschrift für Ethnologie* 32, 87–104.
- Scorza D (1974). 'Sentence structure of the Au language.' *Workpapers in Papua New Guinea Languages* 1, 165–246.
- Scorza D (1985). 'A sketch of Au morphology and syntax.' In *Papers in New Guinea Linguistics*, vol. 22. Canberra, Australia: Pacific Linguistics. 215–273.
- Schmidt W & Spölgen N (1901). 'Beiträge zur Kenntnis der Valman-Sprache.' *Wiener Zeitschrift für die Kunde des Morgenlandes* 15, 335–366.
- Staley W E (1994). 'Theoretical implications of Olo verb reduplication.' *Language and Linguistics in Melanesia* 25, 185–190.
- Staley W E (1996a). 'The multiple processes of Olo verb reduplication.' *Language and Linguistics in Melanesia* 27(2), 147–174.
- Staley W E (1996b). Serial clauses in Olo. Paper presented at the 1996 meeting of the Linguistic Society of Papua New Guinea.
- Vormann F & Scharfenberger W (1914). *Anthropos linguistische bibliothek: internationale sammlung linguistischer monographien: Die Monumbo-Sprache: Grammatik und Wörterverzeichnis*. Vienna: Mechitharistenbuchdruckerei.

Totonacan Languages

C J MacKay and F R Trechsel, Ball State University, Muncie, IN, USA

© 2006 Elsevier Ltd. All rights reserved.

The Totonacan languages are spoken in central Mexico in a region that includes parts of three states: southern Hidalgo, northern Puebla, and northwestern Veracruz (see **Figures 1** and **2**). Although proposals have sometimes been made to relate the Totonacan languages to Mayan, Mixe-Zoquean, and other languages in Mesoamerica (McQuown, 1942), these relationships have never been demonstrated. Today, the Totonacan language family is generally regarded as an 'isolate' in the classification of Mesoamerican languages (Suárez, 1983; Campbell, 1997). It is thought that speakers of these languages settled near the Gulf Coast around 800 A.D. Their original homeland is unknown; however, based on ethnohistorical sources and loanwords found in other Mesoamerican languages, it has been proposed that Totonacs may have founded Teotihuacan and moved to their current location following its collapse (Justeson *et al.*, 1985).

Totonacan Language Family

The Totonacan language family is made up of two branches: Totonac, consisting of four languages, with

roughly 220 736 speakers, and Tepehua, consisting of three languages, with approximately 8252 speakers (INEGI–XII Censo General, 2000). Although the Totonac and Tepehua languages are mutually unintelligible today, they share a great deal of vocabulary and exhibit many structural similarities. These similarities indicate that the languages developed, historically, from a common ancestor, Proto-Totonacan. **Figure 3** provides a simplified representation of the relationships of the various languages. As linguistic investigation proceeds, further groupings and subgroupings within the family will undoubtedly emerge.

As illustrated in **Figure 3**, the Totonac branch consists of four languages, referred to here as Misantla, Papantla, Sierra, and Northern:

Misantla Totonac, the southernmost variety, is spoken between the cities of Xalapa and Misantla in Veracruz. Towns where speakers may still be found include Yecuatla (192 speakers), San Marcos Atexquilapan (13), Landero y Coss (61), Chiconquiaco (56), and Jilotepec (11) (INEGI–XII Censo General, 2000). Misantla Totonac is moribund, with few native speakers remaining, all over the age of 45. The largest concentration of speakers is found in Yecuatla, but their number is dwindling rapidly. According to the Mexican Census, 486 individuals spoke Totonac in Yecuatla in 1980; in 2000, only



Figure 1 Mexico (adapted from a map drawn by Ashley Withers).

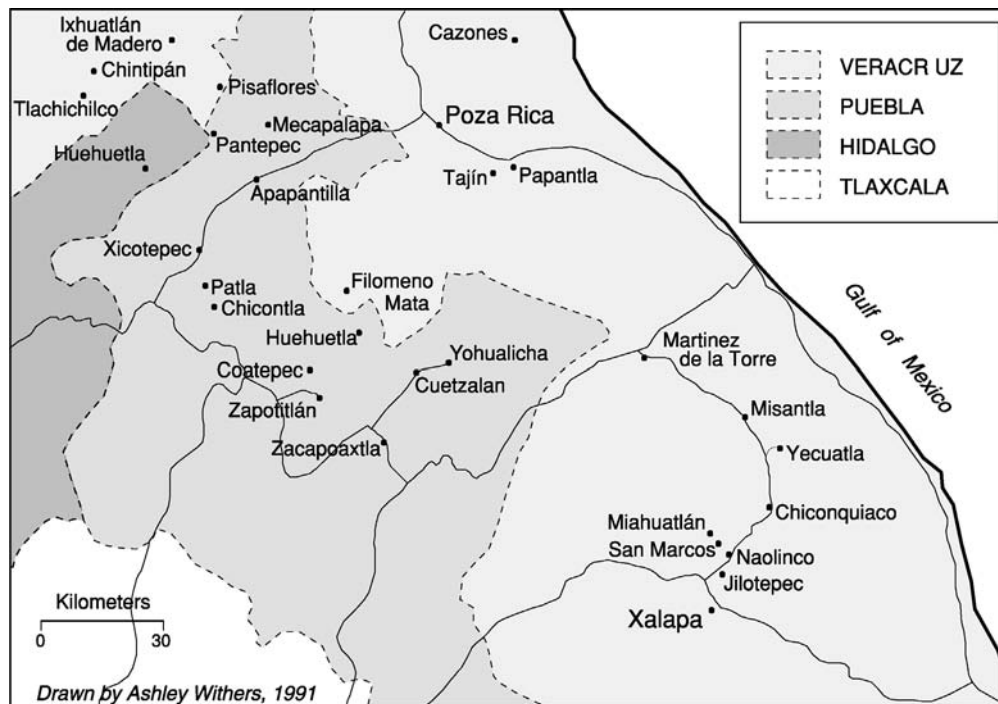


Figure 2 Totonacan language area (adapted from a map drawn by Ashley Withers).

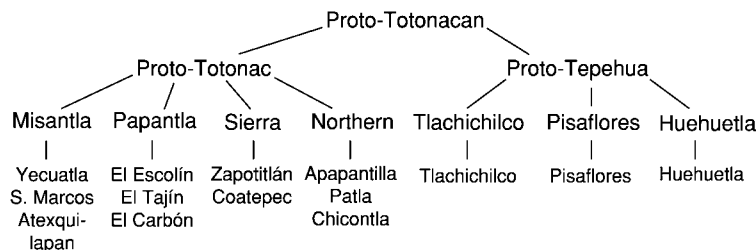


Figure 3 Totonacan language family.

192 speakers remained (INEGI–Censo General, 1980, 2000). Data on Misantla Totonac come from Yecuatla and San Marcos Atexquilapan (MacKay, 1994, 1999; MacKay and Trechsel, 2003, in press).

Papantla Totonac is spoken by roughly 36 000 individuals in and around the city of Papantla, Veracruz. Children are still learning Papantla Totonac, but the language is being used less frequently within the communities. Data on Papantla Totonac come from El Escolín (Aschmann, 1973), Cerro del Carbón (Levy, 1987, 1990), and El Tajín (García Ramos, 2000).

Sierra Totonac is spoken by more than 100 000 people in the Sierra Norte de Puebla and nearby towns in Veracruz. The exact limits of Sierra Totonac and Northern Totonac are still being determined. Children continue to learn Sierra Totonac as their native language, and it is the main language used in many communities. Data on Sierra Totonac come

from Zapotitlán de Méndez, Puebla (Aschmann and Wonderly, 1952; Aschmann, 1962), and Coatepec, Puebla (McQuown, 1990).

Northern Totonac is spoken by roughly 10 000 people in the region surrounding Xicotepec de Juárez, Puebla. It is unclear how many children are learning Northern Totonac; most speakers appear to be middle aged or older. Data on Northern Totonac come from Apapantilla, Puebla (Reid *et al.*, 1968; Reid and Bishop, 1974; Reid, 1991) and Patla and Chicontla, Puebla (Beck, 2004). Beck refers to the variety of Northern Totonac spoken in the latter two communities as Upper Necaxa Totonac.

The Tepehua branch of Totonacan consists of three languages, identified here as Tlachichilco, Pisaflores, and Huehuetla.

Tlachichilco Tepehua is spoken in Tlachichilco, Veracruz, and in the surrounding communities of Chintipán, Tierra Colorada, and Tecomajapa.

According to the 2000 census, there are approximately 2463 speakers in these communities, many of whom are middle aged or older (Watters, 1988: 5). James K. Watters is the only linguist to have conducted research on Tlachichilco Tepehua. His publications include discussions of morphosyntax (1988), phonology (1987), verbal semantics (1996), and second-person laryngealization (1994). Although there is as yet no published lexicon or descriptive grammar of Tlachichilco Tepehua, it is the best documented of all the Tepehua languages.

Pisaflores Tepehua is spoken by roughly 2786 individuals in and around Pisaflores, Veracruz. Tepehua is the main language of this community and children are still learning it as their native language. Carolyn J. MacKay and Frank R. Trechsel have been conducting linguistic research in Pisaflores since 1997 and are working on a description of Pisaflores Tepehua.

Huehuetla Tepehua is spoken in and around the towns of Huehuetla, Hidalgo, and Mecapalapa, Puebla. There are approximately 1649 speakers of Huehuetla Tepehua, all of whom are at least middle aged. Publications on the language include a short sketch of sentence structure, a description of Tepehua numerals, and a preliminary description of verb inflection (Herzog, 1974). The Liga Bíblica Mundial del Hogar published the New Testament in Huehuetla Tepehua in 1976.

Totonacan Phonology

Totonacan languages exhibit three vowels, /a/, /i/, /u/, and a length distinction, contrasting short and long vowels. Some languages, like Northern Totonac, have also developed phonemic /e/ and /o/ (Beck, 2004). Plain and laryngealized variants of both short and

long vowels exist in all Totonacan languages. Whether this distinction is contrastive or predictable has not yet been determined for all varieties. However, all Totonacan languages employ laryngealization to mark second-person subjects.

- (1) Misantla Totonac (MacKay, 1999: 156, 157)

| | |
|-----------------------|----------------|
| [wɨ́f kʌtsɨ́] | ‘you know X’ |
| /wɨ́f kʌtsii/ | |
| [ʔɨ́f kʌtsii] | ‘s/he knows X’ |
| /ut kʌtsii/ | |
| [kinʌ́n kʌ́yʌ́a] | ‘we cut X’ |
| /kinʌ́n kʌ́-yʌ́a-wa/ | |
| [wɨ́ʃin kʌ́yʌ́atʌt] | ‘y’all cut X’ |
| /wɨ́ʃin kʌ́-yʌ́a-tʌt/ | |

Figure 4 presents the consonants that are found in almost all Totonacan languages. In most Totonacan languages, glottal stop is contrastive only in word-final position. However, in Upper Necaxa Totonac, Pisaflores Tepehua, and Huehuetla Tepehua, /ʔ/ has replaced /q/ and therefore occurs in other positions as well.

Consonant alternations to mark degrees of size, force, and intensity have been described in Totonacan. This sound symbolism typically involves the sets of sounds *s/f/ɬ*, *k/q*, and *ts/tʃ* with *ɬ*, *q*, and *tʃ* being the most intense (Bishop, 1984; Levy, 1987; MacKay, 1999; Beck, 2004).

- (2) Misantla Totonac (MacKay, 1999: 114)

| | | |
|-----------|----------|---------------|
| [tsʊtsɨ́] | /tsʊtsū/ | ‘s/he smokes’ |
| [tʃʊtʃɨ́] | /tʃʊtʃū/ | ‘s/he sucks’ |

- (3) Papantla Totonac (Levy, 1987: 115)

| | |
|------|---------------------|
| sukū | ‘small hole’ |
| ɬukū | ‘medium-sized hole’ |
| ɬuqū | ‘large hole’ |

| | Labial | Alveolar | Alveo-palatal | Palatal | Velar | Uvular | Glottal |
|---------------------|--------|---------------------|---------------|---------|-------|--------|---------|
| Plosive | p | t | | | k | q | ʔ |
| Affricate | | ts, tɬ ^a | tʃ | | | | |
| Nasal | m | n | | | | | |
| Fricative | | s, ɬ | ʃ | | | | h |
| Approximant | w | | | y | | | |
| Lateral approximant | | l | | | | | |

^atɬ is found only in Sierra Totonac, Northern Totonac and Papantla Totonac.

Figure 4 Totonacan consonants.

Totonacan Morphology

Totonacan languages exploit a very complex and productive morphology, characterized by a large number of affixes, both prefixes and suffixes, that do most of the work of the grammar. Verbs and nominals are the major word classes.

Nominals

In some languages (e.g., Misantra Totonac and Sierra Totonac), adjectives and nouns do not differ in their inflectional morphology. In others, however (e.g., Papantla Totonac and Upper Necaxa Totonac), nouns and adjectives are distinct. Nominal inflectional morphology is relatively simple. Nominals are optionally marked for plurality, and in possessive constructions are also marked for person (and sometimes number) of possessors.

- (4) Misantra Totonac (MacKay, 1999: 349)
- | | | |
|--------------------|------------------------|--------------|
| [kíptʃík] | 1POSS-house | 'my house' |
| /kin-tʃík/ | | |
| [kíptʃíkʲɪn] | 1POSS-house-PL | 'my houses' |
| /kin-tʃík-VVn/ | | |
| [kíptʃíkʲɪn] | 1POSS-house-POSS.PL | 'our house' |
| /kin-tʃík-kan/ | | |
| [kíptʃíkʲɪnkʲɪn] | 1POSS-house-PL-POSS.PL | 'our houses' |
| /kin-tʃík-VVn-kan/ | | |

Numerals

The Totonacan numerical system, like many others in Mesoamerica, is vigesimal. In many of the Totonacan languages, the numerical system is being replaced by the Spanish one.

- (5) Misantra Totonac (MacKay, 1999: 393, 394)
- | | |
|---------------------|------------------------|
| [puʃúmpuʃúmpuʃún] | 'sixty (20 + 20 + 20)' |
| /puʃum-puʃum-puʃum/ | |
| [tutún puʃún] | 'sixty (3 × 20)' |
| /tutun puʃum/ | |

Body Part Prefixes

Body Part Prefixes occur on both nominal and verb stems, but are most productive on verbs. They usually denote either the body part affected by the action of the verb or a spatial relationship ('in front of,' 'behind,' 'beside,' 'above,' etc.).

- (6) Misantra Totonac (MacKay, 1999: 230)
- | | |
|--------------------------------------|--|
| [míntaqʷaqʷɪt] | |
| /min-ta-qʷaqʷ-nuʷ-Vt/ | |
| 2POSS-INCHOATIVE-ear rel.-inside-NOM | |
| 'your earring' | |

Verbal Inflection

Totonacan verbal morphology is characterized by a layering of derivational and inflectional affixes. Verbs

may be inherently stative, intransitive, transitive, and, in some languages, ditransitive. In all languages, the verbal inflectional system distinguishes two aspectual categories (perfective and imperfective); two tense categories (past and nonpast); and two mood categories (realis and irrealis). In many, but not all, Totonacan languages, the inflectional system also marks categories of future tense and/or perfect aspect. The exact distribution of these latter categories in the family has yet to be determined.

In addition, in all Totonacan languages, verbal inflectional affixes mark categories of person and number of both subjects and objects. For the most part, inflectional affixes are transparent in the sense that they can be easily isolated and their semantic contribution is clear. In transitive sentences involving two nonthird-person arguments, however, certain contrasts are neutralized. In all languages except Sierra Totonac, combinations of a second-person subject and a first-person object, where either or both are plural, are expressed by means of reciprocal verbs with first person inclusive plural subjects. Sentences like the following, from Pisaflores Tepehua, are systematically ambiguous:

- (7) Pisaflores Tepehua (MacKay and Trechsel, 2003: 295)
- | | |
|--|--|
| [kiláaláʔtsínáaw] | |
| /kin-laa-laʔtsín-yaa-wi/ | |
| 1OBJ = RECIP-see.X-IMPERF-1SUBJ.PL | |
| 'You (sg.) see us,' 'You (pl.) see me,' 'You (pl.) see us' | |

A similar ambiguity emerges in sentences in which a first person subject acts on a second person object and, again, one or both are plural. In the Tepehua languages, these combinations are expressed by means of reciprocal verbs with first person exclusive plural subjects. The example in (8) is four-ways ambiguous:

- (8) Pisaflores Tepehua (MacKay and Trechsel, 2003: 297)
- | | |
|--|--|
| [ʔikláláʔtsínáaw] | |
| /ik-laa-laʔtsín-yaa-wi/ | |
| 1SUBJ-RECIP-see.X-IMPERF-1SUBJ.PL | |
| 'I see you (pl.),' 'We see you (sg.),' 'We see you (pl.),' | |
| 'We (excl.) see each other' | |

In contrast, the Totonac languages, with the exception of Sierra Totonac, use reciprocal verbs in 2SUBJ > 1OBJ contexts, but not in 1SUBJ > 2OBJ contexts. Nevertheless, all Totonac languages employ a single verb form to express combinations of first person subject and second person object where one or both are plural. Ambiguities of the sort illustrated in (7) and (8) are pervasive throughout the family.

Verbal Derivation

Totonacan languages exhibit a rich inventory of derivational affixes that affect the valence of both

transitive and intransitive verbs. The most productive are a causative affix, /maa-/ 'CAUS,' and several applicative affixes that license arguments interpreted as beneficiary, recipient/goal, instrumental, comitative, and others. In many languages, applicative affixes are the only means available for expressing arguments with these semantic roles.

- (9) Misantla Totonac (MacKay, 1999: 274)
 [ʔikliləqənāan
 /ik-lil-laqan-yaa-na
 1SUBJ-INST-see X-IMPERF-2OBJ
 kililəqtʃaqáatáayat]
 kin-lil-laq = tʃaqaa-taaya-Vt/
 1POSS-INST-eye rel.-upright-NOM
 'I see you with my glasses'

On transitive verbs, causative and applicative affixes yield ditransitive verbs with two nonoblique objects. There is variation within the family concerning the treatment of these objects. At one extreme are languages like Misantla Totonac in which either or both of the objects may control overt object agreement. Sentences like (10) are systematically ambiguous in this language:

- (10) Misantla Totonac (MacKay, 1999: 190)
 [ʔiklāamakaʔikj
 /ik-laa-maka-ikj-wa
 1SUBJ-3OBJ.PL-hand rel.-give X to Y-1SUBJ.PL
 hōnlībru]
 hun-libru/
 DET-book
 'we handed them the book,' 'we handed him/
 her the books'

At the other extreme are languages like Papantla Totonac (Levy, 2000) in which only one of the two objects may control agreement.

- (11) Papantla Totonac (Levy, 2000: 5)
 ka:ma:xi':-lh lakcumajān kin-qa'wasa
 OBJ.PL-give-PFV girls 1POSS-son
 'I gave my son to the girls' / *'I gave the girls to
 my son'

Between these extremes are several intermediate types in which possibilities of double object marking are constrained by person and number features of the objects.

Totonacan Syntax

Word order in Totonacan languages is extremely flexible and almost any order is acceptable. In unmarked cases, word order is verb initial, and frequently VSO. Subjects may precede the verb for pragmatic effects associated with focus or topicalization.

Coordination and subordination are not explicitly marked and verbs in both clauses exhibit finite verbal morphology.

- (12) Misantla Totonac (MacKay and Trechsel, in press)
 [lakaa ʔiknispāa hōn tʃikjku?
 /lakaa ik-nispaa hun tʃikjku?
 NEG 1SUBJ-know.X DET man
 hōn tiyūut lāatāt]
 hun tiyūut laa-min(tān)-ti/
 DET who COM-come-2PERF
 'I don't know the man you came with.'

Bibliography

- Aschmann H P (1962). *Vocabulario totonaco de la Sierra. Serie de Vocabularios Indígenas 'Mariano Silva y Aceves'.* Núm. 7. México, DF: Instituto Lingüístico de Verano.
 Aschmann H P (1973). *Diccionario totonaco de Papantla. Serie de Vocabularios y Diccionarios Indígenas 'Mariano Silva y Aceves'.* Núm. 16. México, DF: Instituto Lingüístico de Verano.
 Aschmann H P & Wonderly W (1952). 'Affixes and implicit categories in Totonac verb inflection.' *International Journal of American Linguistics* 18, 130–135.
 Beck D (2001). 'Person-hierarchies and the origin of asymmetries in Totonac verbal paradigms.' *Linguistica Atlantica* 23, 1–33.
 Beck D (2004). *Upper Necaxa Totonac*. Muenchen: Lincom Europa.
 Bishop R G (1984). 'Consonant play in lexical sets in Northern Totonac.' *S. I. L.-Mexico Workpapers* 5, 24–31.
 Campbell L (1997). *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press.
 García Ramos C (2000). *Vocabulario bilingüe Totonaco-Castellano*. Jalapa: Ediciones Cultura de Veracruz.
 Herzog D L (1974). 'Person, number, and tense in the Tepehua verb.' *S. I. L.-Mexico Workpapers* 1, 45–52.
 Instituto Nacional de Estadística, Geografía e Información – INEGI (1980). *X Censo General de Población y Vivienda, 1980*. México, DF: Secretaría de Programación y Presupuesto.
 Instituto Nacional de Estadística, Geografía e Información – INEGI (2000). *XII Censo General de Población y Vivienda, 2000*. México, DF: Secretaría de Programación y Presupuesto.
 Justeson J S, Norman W, Campbell L & Kaufman T (1985). *The foreign impact on lowland Mayan languages and script* (Publication No. 53). New Orleans: Middle American Research Institute, Tulane University.
 Levy P (1987). *Fonología del totonaco de Papantla, Colección Lingüística Indígena No. 3. Instituto de Investigaciones Filológicas Veracruz*. México, DF: Universidad Nacional Autónoma de México.
 Levy P (1990). *Totonaco de Papantla, Veracruz*. Archivo de lenguas indígenas de México. México, DF: El Colegio de México.

- Levy P (2000). 'El aplicativo dativo/benefactivo en totonaco de Papantla.' *Memorias del VI encuentro de lingüística en el noroeste*. Sonora: UNISON.
- Liga Bíblica Mundial del Hogar (1976). *El Nuevo Testamento en el idioma tepehua de Huehuetla Hidalgo*. México, DF: Sociedades Bíblicas Unidas.
- MacKay C J (1991). A grammar of Misantla Totonac. Ph.D. diss., The University of Texas, Austin.
- MacKay C J (1994). 'A sketch of Misantla Totonac phonology.' *International Journal of American Linguistics* 60, 369–419.
- MacKay C J (1999). *A grammar of Misantla Totonac*. Salt Lake City: University of Utah Press.
- MacKay C J & Trechsel F R (2003). 'Reciprocal /laa-/ in Totonacan' *International Journal of American Linguistics* 63, 275–306.
- MacKay, C J & Trechsel, F R (in press). *Totonaco de Misantla, Veracruz*. Archivo de Lenguas Indígenas de México. México, DF: El Colegio de México.
- McQuown N A (1942). 'Una posible síntesis lingüística Macro-Mayence.' In *Mayas y Olmecas*. Tuxtla Gutierrez, Chiapas: Sociedad Mexicana de Antropología.
- McQuown N A (1990). *Gramática de la lengua totonaca*. Colección Lingüística Indígena No. 4. Instituto de Investigaciones Filológicas. México, DF: Universidad Nacional Autónoma de México.
- Reid A (1991). *Gramática totonaca de Xicotepec de Juárez, Puebla*. Serie de Gramáticas de Lenguas Indígenas de México. Núm. 8. México, DF: Instituto Lingüístico de Verano.
- Reid A & Bishop R G (1974). *Diccionario totonaco de Xicotepec de Juárez, Puebla: Totonaco-Castellano and Castellano-Totonaco*. Serie de Vocabularios y Dicciones Indígenas 'Mariano Silva y Aceves'. Núm. 17. México, DF: Instituto Lingüístico de Verano.
- Reid A, Bishop R G, Button E M & Longacre R E (1968). *Totonac: from clause to discourse*. S. I. L. Publications in Linguistics and Related Fields 17. Norman, Oklahoma: Summer Institute of Linguistics of the University of Oklahoma.
- Suárez J A (1983). *The Mesoamerican Indian languages*. Cambridge: Cambridge University Press.
- Watters J K (1987). 'Underspecification, multiple tiers, and Tepehua phonology.' In Bosch A, Need B & Schiller E (eds.) *Parasession on autosegmental and metrical phonology*. Chicago: Chicago Linguistic Society. 389–402.
- Watters J K (1988). Topics in Tepehua grammar. Ph.D. diss., University of California, Berkeley.
- Watters J K (1994). 'Forma y función de la morfología de segunda persona en tepehua.' In MacKay C J & Vázquez V (eds.) *Investigaciones lingüísticas en Mesoamérica*. México, DF: Universidad Nacional Autónoma de México. 211–226.
- Watters J K (1996). 'Frames and the semantics of applicatives in Tepehua.' In Casad E H (ed.) *Cognitive linguistics in the Redwoods: the expansion of a new paradigm in linguistics*. Berlin: Mouton de Gruyter.

Trans New Guinea Languages

A Pawley, Australian National University,
Canberra, Australia

© 2006 Elsevier Ltd. All rights reserved.

The Trans New Guinea Family

Comprising upward of 400 languages, Trans New Guinea (TNG) is the third largest family in the world in number of languages, behind Austronesian and Niger-Congo and ahead of Indo-European. TNG is the predominant family on the large island of New Guinea, a region of spectacular linguistic diversity that contains some 18 families that are not demonstrably related (see **Papuan Languages** and **Austronesian Languages**). TNG languages are spoken continuously along the 2000-km mountain chain that runs along the center of New Guinea as far west as the Bird's Head, and they also are used in several parts of the lowlands. At least a dozen TNG languages are also present on Timor, Alor, and Pantar Islands in East Nusantara.

About 3 million people speak TNG languages. Yet, most of the languages have fewer than 5000 speakers. Their small size reflects the difficult terrain of New Guinea in combination with extreme political fragmentation; peoples were traditionally subsistence farmers or foragers, and until colonial times political groups seldom exceeded a few hundred people. The largest TNG language communities are Enga (about 200 000) and Medlpa (Melpa; 150 000) in the highlands of Papua New Guinea, and Western Dani (150 000) and Lower Grand Valley Dani (130 000) in the highlands of West Papua (Irian Jaya).

Until the late 19th century, the TNG languages of New Guinea were completely unknown to linguists, and most remained unrecorded until after World War II. Since then, linguists from various parts of the world have done descriptive and comparative work on TNG languages. Although most are still only documented (at best) by grammatical sketches and word lists, there are quite detailed published grammatical descriptions of perhaps 50 to 70 TNG languages. Reasonably

good dictionaries exist for about 20 TNG languages. Excellent introductory overviews are given in Foley (1986, 2000). The atlas of Wurm and Hattori (1981–83) contains detailed maps, and Carrington's work (1996) is a near-exhaustive bibliography. Languages for which there are good grammars include Korafe of the Binandere group (Farr, 1999), Grand Valley Dani (Bromley, 1981), Hua (a dialect of Yagaria) of the Gorokan group (Haiman, 1980), and Eipo (Eipomek) of the Mek group (Heeschen, 1998).

History of the Trans New Guinea Hypothesis

The hypothesis that there is a large TNG family was proposed about 1970 by linguists at the Australian National University, mainly on the basis of typological resemblances and a handful of widespread putative cognates (McElhanon and Voorhoeve, 1970; Wurm, 1975). However, critics argued that the hypothesis was based on unreliable methods and that the evidence was unconvincing. Percentages of resemblant basic vocabulary forms shared by languages belonging to distant branches of TNG are very low, in the range of 3–7%, and in a region where there has been extensive lexical diffusion for millennia, this level of agreement could be due to borrowing and chance. Recently, linguists have applied more classical comparative methods and have found evidence that strongly supports a modified version of the TNG hypothesis (Pawley, 1995; Ross, 1995; Pawley, 1998, 2001, in press; Ross, in press).

The main grounds for considering TNG to be a language family are (1) systematic form-meaning correspondences in the independent personal pronouns, permitting reconstruction of virtually a complete paradigm (Table 1); (2) some 200 putative cognate sets (nearly from 'basic vocabulary') being represented in two or more major subgroups (Table 2); (3) a body of regular sound correspondences for a small sample of languages belonging to eight different subgroups, which has allowed a good part of the Proto TNG sound system to be reconstructed (Table 3); and (4) resemblances in certain other grammatical paradigms, chiefly the form of verbal suffixes marking

person-number of subject. In addition, the distribution of certain striking structural features, such as switch-reference morphology on verbs, has been shown to correlate rather closely with the distribution of TNG languages.

TNG seems to have had a simple syllable structure, with syllables of the shape (C)V and (word-finally) CVC.

Subgroups of Trans New Guinea

More than 30 subgroups are recognized that have not been assigned to any larger grouping within Trans New Guinea. Much of the evidence for these groups is based on innovations in the personal pronouns (Ross, in press). The following is a selection of the more important subgroups.

Madang (Madang-Adelbert Range) is by far the largest well-defined subgroup of TNG, with about 100 members (*see Madang Languages*). It occupies the central two-thirds of Madang Province from the coast to the Bismark and Schrader Ranges. Huon-Finisterre contains about 70 languages spoken on the Huon Peninsula and in the Finisterre and Saruwagi Ranges in Morobe and Madang Provinces.

Table 2 Some cognate sets of the Trans New Guinea family

| | 'breast' | 'eat' | 'louse' | 'name' |
|--------------------------------|----------|-------|---------|--------|
| Proto TNG | *amu | *na- | *niman | *imbi |
| Asmat (Irian Jaya) | | na- | | yipi |
| Kiwai (SW coast, PNG) | amo | | nimo | |
| Kewa (W. Highlands, PNG) | | na- | | ibi |
| Kuman (C. Highlands, PNG) | aemu | | numan | |
| Kube (Morobe, PNG) | namu | ne- | imiŋ | |
| Katiati (Madang Province, PNG) | ama | | ñima | nimbi |
| Aomie (Central Province, PNG) | ame | | ume | ihe |

Table 3 Proto TNG segmental phonemes (Minimal set)

| | Bilabial | Apical | Palatal | Velar |
|-------------------------|----------|---------|---------|-------|
| Consonants | | | | |
| oral obstruents | p | t s | | k |
| prenasalized obstruents | mb | nd | nj | ŋg |
| nasals | m | n | | ŋ |
| lateral glide | w | l | y | |
| Vowels | front | central | back | |
| high | i | | u | |
| mid | e | | o | |
| low | | a | | |

Table 1 Proto TNG free pronouns

| | 1st person | 2nd person | 3rd person |
|------------------|------------|------------|------------|
| sing. | na | ŋga | [y]a, ua |
| pl. (i-graduate) | ni | ŋgi, ki | i |
| (u-graduate) | nu | | |
| pl. | nja | | |

Chimbu-Wahgi (Chimbu) is centered east and south of Mt. Hagen, in the Wahgi, Nebilyer, and Kaugel Valleys and extends north of the Sepik-Wahgi Divide into the Jimi Valley. It contains perhaps 12 languages, although the situation is complicated by extensive dialect chaining. The best-known members are probably Kuman (Chimbu), Middle Wahgi, Sinasina, and Medlpa (Melpa). Engan is a well-defined group consisting of several languages spoken over a wide area to the west of Mt. Hagen. There is a northern subgroup that includes Enga, Ipili, Iniai (Bisorio), and Lembena and a southern subgroup that includes Sau (Samberigi), Huli, Mendi (Angal), and Kewa.

The Kainantu and Goroka groups occupy contiguous parts of Eastern Highlands Province. Each consists of a half a dozen or so languages, some with diverse dialects. Together they probably form a single higher-order larger subgroup, Kainantu-Goroka. The Angan group of about 12 languages occupies considerable areas of Morobe and Gulf Provinces and extends into the Eastern Highlands province. Southeast New Guinea contains the Dagan, Mailuan, Yareba, Manubaran, Kwalean, and Koiari groups, which all replace Proto TNG **ngi* '2 PL' by **ya*, as well as the Binandere and Goilalan groups.

The Ok group comprises about about 10 languages spoken in the central ranges around the West Papua-Papua New Guinea border, including the Star Mountains, and the Thurnwald and Victor Emmanuel Ranges. The Awyu-Dumut and Asmat-Kamoro groups occupy the lowlands to the southwest of this area, in West Papua. The Dani languages spoken in and around the Baliem Valley and the Wissel Lakes languages seem to belong together in a Western New Guinea group. The West Bomberai and Timor-Alor-Pantar groups share two probable innovations in pronouns, which suggests that together they may form a West Trans New Guinea group (Figure 1).

Where and When was Proto TNG Spoken?

The largest concentration of established high-order subgroups of TNG lies in the central highlands of Papua New Guinea between the Strickland River and Eastern Highlands. It is safe to say that this was a very early area of TNG expansion and that initial dispersal was mainly along the central cordillera. If we take conventional estimates for the breakup of Indo-European (at least 6000 years ago) and Austronesian (about 5000 years ago) as yardsticks, a date of between 8000 and 12 000 years ago for the breakup of TNG is reasonable, given that lexicostatistical diversity within TNG is far greater than in either Indo-European or Austronesian. It is noteworthy that dates of about 10 000 years ago have been established

for early agriculture, probably based on taro and bananas, in the Upper Wahgi Valley (Denham *et al.*, 2003). It may have been their use of agriculture that enabled speakers of TNG languages to establish permanent settlements along the central highlands of New Guinea as the climate warmed after the last Ice Age.

Structural Characteristics of TNG Languages

Phonology

Many TNG languages have sound systems similar to that posited for Proto TNG, with syllables of the shape (C)V and (word-finally) CVC, five vowels, and series of nasals, oral, and pre-nasalized (or voiceless and voiced) obstruents with contrasts at bilabial, apical, and velar (and sometimes palatal) positions. A number of languages in the central highlands have a contrast between dental, alveo-palatal, and velar laterals, or between resonant and fricative palatals. [t] and [r] are often allophones of the same phoneme. Many TNG languages have word tone or pitch accent (Donohue, 1997).

Grammar and Semantics

The preferred order of constituents in verbal clauses is SOV, but OVS often occurs as a marked structure. Adpositions follow the verb, whereas determiners and possessors follow the noun. Case marking is generally absent or little developed. Most languages organize pronominal affixes to show a nominative-accusative (or dative) contrast. No language is known to have a full ergative-absolutive alignment for verb pronominals.

Generally, a verb root cannot be used as a noun without derivational morphology or vice versa. In some languages verb roots are a small closed class, with between 50 and 150 members. The densest concentration of such languages seems to be in the Chimbu-Wahgi and Kalam-Kobon subgroups. Common nouns are an open class with many subclasses. Minor classes include adjectives, adverbs, and (see below) verbal adjuncts.

TNG languages typically have fairly simple systems of independent pronouns, in some cases distinguishing three persons but with no number contrasts. More complex pronoun systems are constructed by adding number markers for dual and plural. However, there is often a discrepancy between the kinds of distinctions made in independent pronouns and in verbal affixes. For example, Kuman of the Chimbu-Wahgi family has only four independent pronouns – first person singular, first person plural, second person,

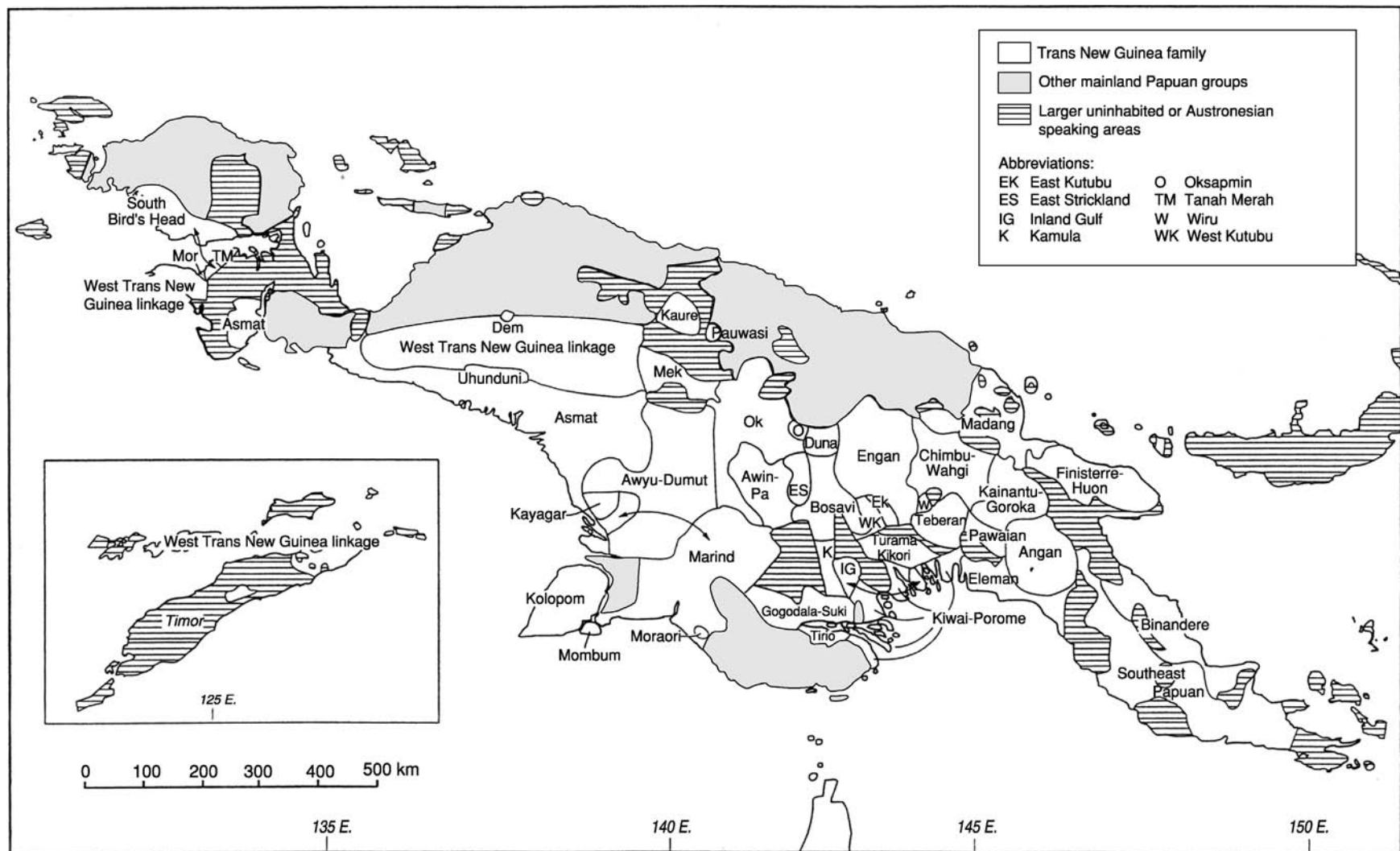


Figure 1 Location of the main subgroups of the Trans New Guinea family.

and third person – but in verbal morphology Kuman makes nine contrasts for the subject: three persons each with singular, dual, and plural.

Morphology is chiefly suffixal. In most languages, nouns carry little morphology. Kinship terms and sometimes part terms often require affixed possessive pronouns. A few languages mark gender contrasts. A good many TNG languages use existential verbs, such as ‘stand’, ‘sit’, ‘lie’ and sometimes, other verbs like ‘hang’, ‘carry’ and ‘come’, as quasi-classifiers of nouns (Lang, 1975), with nouns selecting a verb according to their shape, posture, size, and composition. However, the choice of verb has some flexibility relative to the situation of the referent. Nouns are usually not inflected for number. Certain generic categories are typically expressed by N+N (and occasionally N+N+N) compounds denoting the most salient members of a class (e.g., often ‘people’ is ‘woman-man’, ‘children’ is ‘girl-boy’, and ‘ancestors’ is ‘grandmother-grandfather’).

Most TNG languages distinguish two types of inflected verb, often called ‘final’ and ‘medial.’ Final verbs head the final clause in a sentence and carry suffixes marking absolute tense-aspect-mood and person-number of subject. Medial verbs head nonfinal coordinate-dependent clauses and carry suffixes marking (1) whether the event denoted by the medial verb occurs prior to or simultaneous with that of the final verb and (2) ‘switch reference’ (i.e., whether that verb has the same subject or topic as the next clause; Roberts (1997) surveys the several kinds of switch-reference systems.) In many languages, transitive verbs also carry a pronominal prefix or proclitic marking object agreement. Some languages have causative and applicative affixes that add arguments to the verb. Several Highlands subgroups carry evidential suffixes, indicating whether the clause denotes an event witnessed by the speaker or based on hearsay. In constructions denoting uncontrolled bodily and mental processes (e.g., sweating, sneezing, bleeding, feeling sick), the experiencer is often marked by an object/dative pronoun and is the direct object. A noun denoting the bodily condition is, arguably, the subject or else there is no referential subject. There are usually clauses with nominal predicates denoting class membership and identifying relationships.

All TNG languages make extensive use of at least one of two types of complex (multi-headed) predicates to augment their stock of verbs. First, in verbal adjunct constructions, an inflected verb, usually carrying a rather general meaning, such as ‘make’, ‘hit’ or ‘go’, occurs in partnership with a noninflecting base (the verbal adjunct), which carries a more specific meaning. Verbal adjuncts partner just a small set of verbs. For example, Kalam has a single verb of

sound making and speaking, *ag-*, but has some 30 verbal adjuncts that denote particular kinds of sounds and occur only with *ag-*. Second, in serial verb constructions, two or more bare verb roots occur in sequence to express a tightly integrated sequence of subevents; for example, Kalam: *d ap* (get come) ‘bring’, *d am* (get go) ‘take’, *am d ap* (go get come) ‘fetch’, *d nŋ* (touch perceive) ‘feel’, *ñb nŋ* (eat perceive) ‘taste’. Many languages have a looser type of serial verb construction – narrative serialization – which allows a streamlined, formulaic representation of episodes in which the same actor performs a familiar sequence of actions; for example, Kalam: *am alŋaw-kab tk d ap ad ñb-* (go pandanus-nut gather get come cook eat) ‘gather and eat pandanus nuts’.

Generally, little use is made of conjunctions to show sequential, conditional, and causal relations. Speakers commonly use long chains of clauses headed by medial verbs to report a sequence of past events that make up a single complex episode. In narratives, paragraph-like boundaries are frequently marked by head-tail linkage, in which the last clause of the previous sentence is repeated, to begin a new sentence.

Bibliography

- Baak C, Bakker M & van der Meij D (eds.) (1995). *Tales from a concave world. Liber Amoricum Bert Voorhoeve*. Leiden: Leiden University.
- Bromley M (1981). *A grammar of Lower Grand Valley Dani*. Canberra: Pacific Linguistics.
- Carrington L (1996). *A linguistic bibliography of New Guinea*. Canberra: Pacific Linguistics.
- Denham T P, Haberle S G, Lentfer C *et al.* (2003). ‘Origins of agriculture at Kuk Swamp in the Highlands of New Guinea.’ *Science* 201, 189–193.
- Donohue M (1997). ‘Tone systems in New Guinea.’ *Linguistic Typology* 1, 347–386.
- Foley W (1986). *The Papuan languages of New Guinea*. Cambridge University Press.
- Foley W (2000). ‘The languages of New Guinea.’ *Annual Review of Anthropology* 29, 357–404.
- Farr C (1999). *The interface between syntax and discourse in Korafe: a Papuan language of Papua New Guinea*. Canberra: Pacific Linguistics.
- Haiman J (1980). *Hua: a Papuan language of the Eastern Highlands of New Guinea*. Amsterdam: Benjamins.
- Heeschen V (1998). *An ethnographic grammar of the Eipo language spoken in the central mountains of Irian Jaya (West New Guinea), Indonesia*. Berlin: Dietrich Reimer.
- Lang A (1975). *The semantics of classificatory verbs in Enga (and other Papua New Guinea languages): a semantic study*. Canberra: Pacific Linguistics.
- McElhanon K & Voorhoeve C L (1970). *The Trans-New Guinea phylum: explorations in deep-level genetic relationships*. Canberra: Pacific Linguistics.

- Pawley A (1995). 'C. L. Voorheve and the Trans New Guinea hypothesis.' In Baak *et al.* (eds.), 83–123.
- Pawley A (1998). 'The Trans New Guinea phylum hypothesis: a reassessment.' In Jelle Miedema J, Ode C, Rien A & Dam C (eds.) *Perspectives on the Bird's Head of Irian Jaya, Indonesia*. Amsterdam: Editions Rodopi. 655–689.
- Pawley A (2001). 'The Proto Trans New Guinea obstruents: arguments from top-down reconstruction.' In Pawley A, Ross M & Tryon D (eds.) *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*. Canberra: Pacific Linguistics. 261–300.
- Pawley A (in press). 'The chequered career of the Trans New Guinea hypothesis. Recent research and its implications.' In Pawley *et al.* (eds.). (in press).
- Pawley A, Attenborough R, Hide R & Golson J (eds.) (in press). *Papuan pasts: studies in the cultural, linguistic and biological history of the Papuan-speaking peoples*. Canberra: Pacific Linguistics.
- Roberts J (1987). *Amele*. London: Croom Helm.
- Roberts J (1997). 'Switch reference in Papua New Guinea: a preliminary survey.' In Pawley A (ed.) *Papers in Papuan linguistics* No. 3. Canberra: Pacific Linguistics. 101–241.
- Ross M (1995). 'The great Papuan pronoun hunt: recalibrating our sights.' In Baak *et al.* (eds.). 139–168.
- Ross M (in press). 'Pronouns as a preliminary diagnostic for grouping Papuan languages.' In Pawley *et al.* (eds.). (in press).
- Wurm S A (ed.) (1975). *New Guinea area languages. 1: Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics.
- Wurm S A & Hattori S (1981–83). *Language atlas of the Pacific area* (2 vols). Canberra: Australian Academy for the Humanities in collaboration with the Japanese Academy.

Tsotsi Taal

B B Mfenyana, Kagiso, South Africa

© 1994 Elsevier Ltd. All rights reserved.

Boy Faraday, Bitch Never Die, Bra Slim, Ous Kuki (or Sta Kuja), Bro' Don, Oom Sirra, Bra Terror, Zorro – these names are all part of Tsotsi culture and language, popularly known as Isi-Tsotsi (thug-speak). Like British, American, and other types of slang, black South African taal/lingo has a checked, colorful history stretching back to 1930 and beyond. Other names used for this street variety are Mensetaal ('the language of the people'), *Flytaal*, *Iscamtho* or *Isijita* (lit. the language of the jits or jitas: young townies; Mfenyana 1977).

This witty, controversial, and evanescent argot, whose words are adopted and discarded almost at will, sprouted in the dusty streets of Sophiatown, Alexandra, and Soweto (all urban townships or slums located around Jozi/Johannesburg), and quickly spread to Langa (Cape Town), New Brighton (Port Elizabeth), Duncan Village (East London, SA), Marabastad and Mamelodi (Pretoria), Umlazi (Durban), etc. Over time, and because each province or area is dominated by one or another African language (Sotho, Xhosa, Zulu, Afrikaans, or Tsonga), stylistic, tonal, and vocabulary differences began to emerge among the many types of Tsotsi Taal.

Predictably enough, the South African public is not yet agreed on who or what exactly 'a tsotsi' is:

M. J. H. Mfusi asserts that '...tsotsis were part of the ethnically mixed society of the locations, and among themselves spoke the Afrikaans dialect (flytaal or mensetaal)' (1992: 46). Their lifestyle revolved, as it still does today, around flashy/American clothes, shoes, hats, and motorcars (Glaser: 1992). It must be stressed that, at first, tsotsis did not use much violence to achieve their ends: they relied on their wits, speed, brute strength/size, number of followers, or pure luck. As conditions in the locations and villages worsened (1940–60), tsotsis turned increasingly to rape, armed robbery, and even murder to maintain themselves and their women (called noasias or ootsotsikazi: Town Xhosa). The tsotsi label broadened to include all urban criminals (confidence tricksters and even scholars, known after 1976 as 'comrade-tsotsis' or *com-tsotsi's*; Mfusi 1992: 46).

Thus, by 1990 the label tsotsi could justifiably be applied to any wayward, unreliable, or 'clever' person, young or old. The key issue then becomes identification of Tsotsi Taal speakers. Ordinary people, active and 'retired' crooks, journalists (like the late Casey 'Kid' Motsisi, Can Temba, and currently Bra El Makhaya of the *Sowetan* newspaper), and Bra Obed Musi (City Press); musicians like Ray 'Zwakala Nganeno' (Come nearer) Phiri and Brenda 'Weekend Special' Fassie; poets of the calibre of Bro' Don 'Zirga Special' Mattera and Sipho Sepamla. All these people have one thing in common: they are regular and enthusiastic, creative, and unapologetic users of some form of Tsotsi Taal.

Consider a few, brief examples:

- a. Greetings: *hella*, *heit*, *heita*, *heitadaa* (i.e., 'hello there'; Afrikaans *daar* = 'there'), and most recently *Hola-hola*.
- b. Parting shots: *skhuvet under die corset*, *sweet*, *sharp*, *grand*, and *mojo/moja*.
- c. Nicknames and ways of expressing respect: *Ta Ben* (from Afr *Boeta Ben*), *'Sta May* ('Sister May'), *Ma-Ben-za*, etc.
- d. Expletives (cannot be excluded: this is the language of the 'swearing class!'): *donder* = lit. 'thunder,' means 'beat up' (thus Town Xhosa *uku-donora*, 'to assault'); *foetsek!* = 'go away' (thus Town Xhosa *uku-futsheka*); *fokof* = 'get out of here,' 'go away,' etc.

In brief, the lexicon of Tsotsi/*flaaitaal* covers a whole range of activities and phenomena: food, drinks, women, police, whites, jail, cigarettes, drugs, love, stealing, and dying.

The importance of this argot lies in its potential as a racial, socioeconomic, age, and gender leveller: because it brings black and white, rich and poor,

educated and unlettered, young and old, male and female together in a way no foreign language ever could. With proper recognition, care, documentation, and development, Tsotsi Taal could become South Africa's national language.

Laat ek nou die mova's skep, bricates. Ons sight mekaar burro 'Sta Pallie se cook-dla, late bells. Heitada, hola-hola, is dolly my ma se kind!

(Let me now take my leave, brothers. We'll meet each other at Sister Palesa's shebeen/tavern, after hours. OK, goodbye, all's well my mother's child!).

Bibliography

- Glaser C (1992). The mark of Zorro: Sexuality and gender relations in the Tsotsi sub-culture on the Witwatersrand. *Journal of African Studies* 51, 1.
- Mfenyana B B (1977). Isi Khumsha nesi Tsotsi: The sociolinguistics of school and town sintu. (Master's dissertation, Boston University).
- Mfusi M J H (1992). Soweto Zulu slang: A sociolinguistic study of an urban vernacular. *English Usage in Southern Africa* 23.

Tucanoan Languages

J Barnes, SIL International, Bogota, Colombia

© 2006 Elsevier Ltd. All rights reserved.

Overview

Classification

The Tucanoan languages of Colombia, Peru, Ecuador, and Brazil fall into two major groupings: Western Tucanoan (WT) and Eastern Tucanoan (ET). In the Eastern area, there are two languages that are very different from the other languages: Cubeo (Cub.) and Retuarā/Tanimuca (Tanimuca-Retuarā; Ret.). Waltz and Wheeler (1972: 128–129), recognizing this difference, chose to put Cub. in a category by itself, calling it 'Middle Tucanoan.' (In other publications, the term 'Central Tucanoan' is used, but is obviously not intended as a geographic term as Cub. is on the northern edge of the other ET languages.) Waltz and Wheeler, in their Proto-Tucanoan studies, referring to the differences between Cub. and the other ET languages, said, "The data indicates a possible break between Cub. and Western Tucanoan some time later than that between the Eastern Tucanoan groups and Western Tucanoan." (1972: 128). Ret., which is spoken south of the main ET area, was not included in

the study by Waltz and Wheeler. Strom commented (1992: 1) that there appears to have been considerable influence from the Yucuna language on Ret. Ramirez (1997: 17) classified Ret. as ET, but puts it in a subcategory of its own. Malone (1986: Sect. 9.1) said: "Cubeo and Retuarā group closest to each other, which is shown by innovations for the protoconsonants *b (in the environment $_V_1CV_2$, where C is bilabial [bil]), *s, *w, *y. Cub. tends to have more innovations in common with WT languages (*b/ $_V_1C_{+bil}V_2$, *d/ $_i$, $_Vs$, *y/ $_Vd$), but also has some in common with ET languages (*d, *w); Ret. tends to have more innovations in common with ET languages (*d, *w), but also has some in common with WT languages (*b/ $_V_1C_{+bil}V_2$, *d/ $_Vs$). I consider these two languages to be Middle Tucanoan."

In Table 1, linguistically similar languages are grouped together, based on their development from Proto-Tucanoan (Malone, 1986: Sect. 9.1 & 9.2, and I have added Pisamira [Pis.] to her data). Barasano (Barasana) and Taiwano are listed as one (Bar.), as the only major difference between the two is in pitch-stress (Jones and Jones, 1991: 2). Retuarā and Tanimuca are listed as one, (Ret.), as they differ mainly in a few lexical items (Strom, 1992: 1). Population

Table 1 The Tucanoan language family and population figures

| <i>Languages and groupings</i> | <i>Population figure</i> | <i>Abbreviations</i> |
|--------------------------------|--------------------------|----------------------|
| EASTERN TUCANOAN Group #1 | | |
| Wanano | 1560 | Wan. |
| Piratapuyo | 1070 | Pir. |
| Tucano | 5000 | Tuc. |
| Pisamira | 46 | Pis. |
| Tuyuca | 815 | Tuy. |
| Yurutí | 850 | Yur. |
| Waimajã/Bará | 700 | Wai./Bar. |
| Carapana | 650 | Car. |
| Tatuyo | 350 | Tat. |
| EASTERN TUCANOAN Group #2 | | |
| Siriano | 310 | Sir. |
| Desano | 1760 | Des. |
| Macuna | 550 | Mac. |
| Barasano/Taiwano | 350 | Bas. |
| MIDDLE TUCANOAN | | |
| Cubeo | 6150 | Cub. |
| Retuarã/Tanimuca | 300 | Ret. |
| WESTERN TUCANOAN | | |
| Koreguaje | 2000 | Kor. |
| Secoya | 435 | Sec. |
| Siona | 300 | Sio. |
| Orejón | 300 | Ore. |

figures for Pis. are taken from González (2000: 374), for Wanano (Guanano; Wan.) are from Stenzel (2004: 23), and for Yurutí (Yuruti; Yur.) are from Kinch (personal communication). The rest of the population figures are taken from the Ethnologue.

The last three in the list of ET languages in Group #1 (from Table 1), demonstrate a borrowing from Arawakan languages in the inclusion of a ubiquitous prefix *ka-*, which does not occur in the other Tucanoan languages. See Metzger (1998). Many other well-established borrowings are apparent in the languages, especially from Geral (Nhengatu) and Portuguese. More recent borrowings from Spanish and/or Portuguese, are freely incorporated into the languages and suffixed with the appropriate Tucanoan suffixes.

Location

The Tucanoan languages are spoken in the border regions of Colombia with Brazil, Peru, and Ecuador. The ET languages and Cub. are spoken in the state of Vaupés in Colombia and the state of Amazonas of Brazil. Ret. is spoken in the state of Amazonas in Colombia. See Figure 1. The WT languages are spoken to the west and southwest of the ET languages. Koreguaje (Kor.) is found in the states of Caquetá and Putumayo in Colombia, Secoya (Sec.) near the Putumayo River in Ecuador and Peru, Siona (Sio.) on both sides of the Putumayo River in Colombia and Ecuador, and Orejón (Ore.) south of the Putumayo River in Peru. See Figure 2.

Interrelationship

The ET people mainly define themselves by their language, which is the language of their father. Thus, a Tuyuca (Tuy.) woman, for example, might marry a Siriano (Sir.) man. Each would continue to speak his or her own language. The children learn their mother's language first, and later on switch to speak their father's language. The men in a Sir. village might marry women from different language groups, with the result that the children in the village grow up hearing from two to six languages. For a succinct analysis of the relationships between the different ET groups, see Aikhenvald (2002: 17–28). The Western and Middle Tucanoan groups are permitted, within specific kinship guidelines, to marry those who speak their own language. The Eastern and Middle Tucanoan language groups all share basically the same culture and belief systems, which differ in significant ways from the Western groups.

Language Features

Some of the more interesting features of the Tucanoan languages are the small number of consonants, nasalization and nasal spreading, the system of numerical classifiers, and the use of evidential suffixes on the verb. These, as well as other features, will be discussed below.

Phonological Characteristics

Vowels

The Tucanoan languages, with the exception of Ret., have the vowel inventory as seen in Table 2. Ret. has a five vowel system, lacking the /ɨ/ of the other Tucanoan languages. In all of the Tucanoan languages there are oral vowels in oral syllables and nasal vowels in nasal syllables. The following examples illustrate the six vowel contrasts:

Tuy. (my field data):

- (1) /ba'a/ 'disposable basket'
- /be'e/ 'to split'
- /bi'i/ 'to be similar'
- /bo'o/ 'to desire'
- /bu'u/ 'tucunaré fish'
- /bi'i/ 'pirana'

The following examples illustrate the oral–nasal contrasts:

Bas. (Jones and Jones, 1991: 9; personal communication):

- (2) a/ā wa/wā 'to go'/'to illuminate'
- e/ē eho/ēhō 'type of jungle nut'/'cold (illness)'



Figure 1 Eastern and Middle Tucanoan languages with approximate locations.

| | | |
|-----|---------|---|
| i/ī | i/ī | ‘that (in sight)’/3SING.MASC (PRONOUN) |
| o/ō | oha/ōhā | ‘to enter woods’/‘to untie’ |
| u/ū | udi/ūdī | ‘to inhale’/‘be similar to’ |
| i/ī | ihi/īhī | ‘hereditary chief of a sib’/‘to burn (fire)’ |

The vowel /ɪ/ is a high central unrounded vowel, which is realized phonetically as a back unrounded

vowel in certain environments in Carapana (Car.) and Tatuyo (Tat.) (Gómez-Imbert, 2000: 326 and Barnes field data).

Vowel Sequences Sequences of two contiguous vowels within a morpheme reveal interesting patterns of symmetry. See the description of Sir. (Nagler and Brandrup 1979: 120–122), and also of Pis. (González

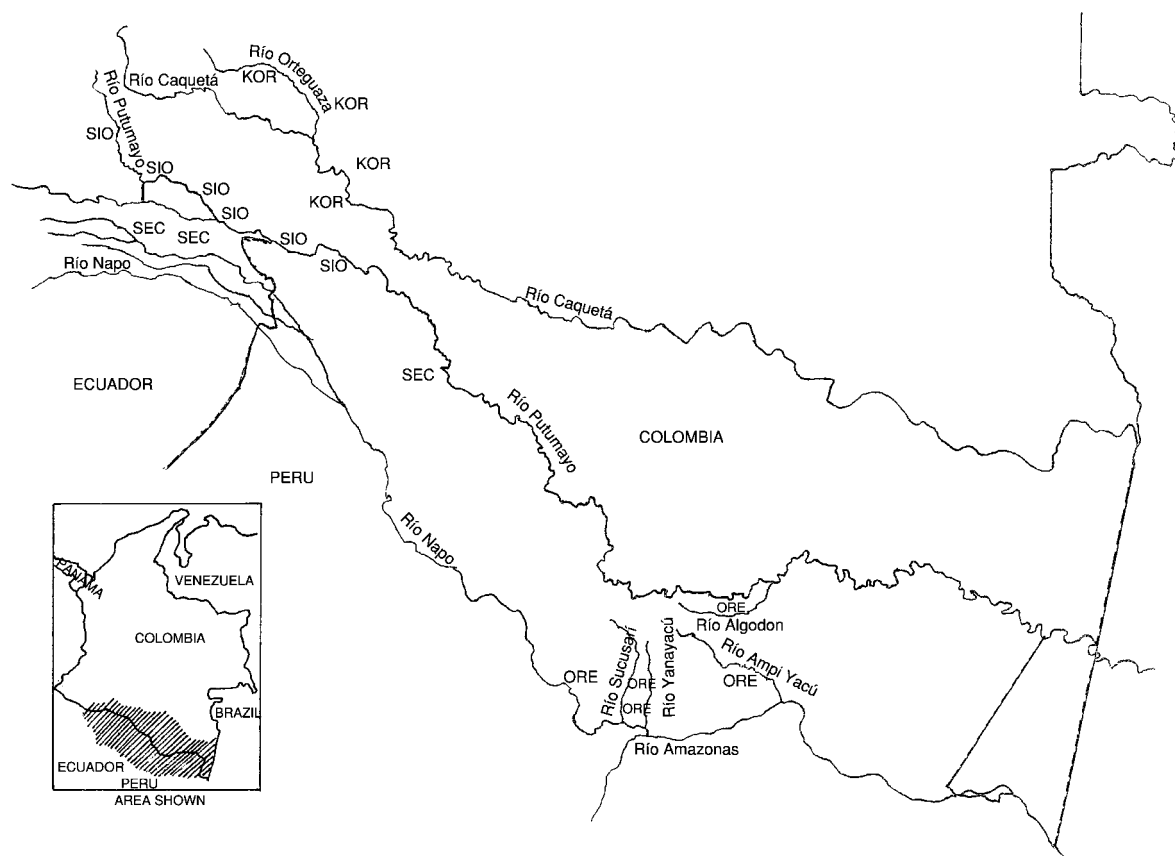


Figure 2 Western Tucanoan languages with approximate locations.

Table 2 Vowel inventory

| Type | Front | Central | Back |
|------|----------|----------|----------|
| High | <i>i</i> | <i>i</i> | <i>u</i> |
| Low | <i>e</i> | <i>a</i> | <i>o</i> |

2000: 380–381), in which she explained that /i/ and /i:/ may precede low vowels in the sequences iV and iV; /u/ may precede the three vowels farthest from itself; /e/ and /o/ may precede other low vowels; and /a/ may precede all three high vowels.

Consonants

The inventory of consonants in the majority of the Tucanoan languages, charted according to their phonemic characteristics, is shown in Table 3.

For a functional analysis of the consonants in four of the ET languages, see Gómez-Imbert (2000: 327–328).

In the ET and Middle Tucanoan languages, the voiced consonants and /h/ have nasal variants in nasal morphemes.

All 12 of the consonants in Table 3 are found in Desano (Des.) and Pir. The following languages have

Table 3 Consonant inventory

| | Bilabial | Alveolar | Palatal | Velar | Glottal |
|----------------------|----------|----------|----------|----------|---------|
| Voiceless stops | <i>p</i> | <i>t</i> | | <i>k</i> | ʔ |
| Voiced stops | <i>b</i> | <i>d</i> | | <i>g</i> | |
| Voiced flap | | <i>r</i> | | | |
| Voiceless sibilant | | <i>s</i> | | | |
| Voiceless semivowels | | | <i>h</i> | | |
| Voiced semivowels | <i>w</i> | | <i>j</i> | | |

all but /ʔ/: Car., Pis., Sir., Tuy., and Yur. The voiceless affricate /tʃ/ is used in Pis. rather than the voiceless sibilant /s/. Waimajã/Bará (Waimaha; Wai.) and Tat. lack both /ʔ/ and /s/. Bas. and Macuna (Mac.) lack /ʔ/ and /p/.

Cub. lacks /ʔ/ as well as /g/, and has the voiceless affricate /tʃ/ rather than /s/. Ret. lacks /j/ and /g/, but has the palatal stop /dj/.

The speakers of languages that lack /p/ or /s/ employ those consonants both in loanwords and when speaking other Tucanoan languages.

Two Western Tucanoan languages, Kor. and Sio., have 24 and 18 consonants respectively. Kor. has no voiced stops, but rather has bilabial and alveolar nasals at those points of articulation, and has /ɲ/ rather than /j/. It includes /w/ in addition to /w/, three voiceless aspirated stops, two voiceless nasals, the voiced affricate /dʒ/, and six labialized consonants. Of the typical 12 Tucanoan phonemes, Sio. lacks /r/, but includes a ‘soft’ voiceless fricative written as /z/, plus the two nasals /m, n/, the voiceless affricate /tʃ/, and three labialized consonants /kʷ, gʷ, hʷ/.

Sec., as with Sio., lacks /r/, but has the same ‘soft’ voiceless fricative written as /z/, plus /kʷ/ rather than /g/, and /m/ rather than /b/, totaling 12.

Ore. lacks /w/ and /r/ ([r] is an allophone of /d/), and includes the voiceless affricate /tʃ/. (Adequate information on Ore. is lacking.)

Two ET languages, Tucano (Tuc) and Wan., have 15 and 16 consonants respectively. Wan., in addition to the typical 12 Tucanoan phonemes, includes the voiceless affricate /tʃ/ and three voiceless aspirated stops. For a description of the development of these additional consonants in present-day Wan. from Proto-Tucanoan, see Waltz 2002. In 1967, West and Welch analyzed Tuc. as having the 12 consonants in Table 3. In 2000, having observed a definite change in pronunciation from, and language attitude toward, what was /CV₁hV₁/ to /ChV₁/ (where C is a voiceless stop), they included voiceless aspirated stops in their analysis, bringing the total of Tuc. consonants to 15. Welch and West (personal communication) analyze the phonetic realization [ChV] as /ChV/, for example: /-kho/ ‘large, FEM,’ whereas Ramirez recognized only /CV₁hV₁/: /-koho/ ‘large’ (1997: 216). Ramirez also analyzed /r/ as an allophone of /d/, which results in his defining a total of 11 consonant phonemes for Tuc. (1997: 25). It is to be noted that West and Welch have studied Tuc. in Colombia, and Ramirez in Brazil. The distance between the two locations of study is so great, and contact between the two groups is so rare, that it is not surprising that there are some differences.

Syllable Patterns

The Middle and ET languages have one basic syllable pattern: (C)V. In Des., Piratapuyo (Pir.), Tuc., and Ret. there is an additional syllable pattern, one in which the syllable is closed with a glottal stop: (C)Vʔ. In all of the languages, (C)Vʔ has been analyzed as two syllables: (C)V.V, except that Ramirez analyzed (C)Vʔ as a bimoraic syllable (1997: 53–56). The WT languages have monosyllabic two-vowel clusters, and a syllable can be closed with a glottal stop in Kor., Sec., and Sio.

There is evidence in Pir., Tuc., and Tuy. that an unstressed vowel between consonants /s/ and /t/ is dropping out, creating the consonant cluster /st/.

Pir. (Klumpp and Klumpp, 1973: 116; Waltz, personal communication):

- (3) /biásitu/ [biástu] ‘pot for hot pepper’

Tuc. (Welch and West, personal communication):

- (4) /baʔa-sité/ [baʔasté] ‘eat and scatter (food)’

Tuy. (my field data):

- (5) /paasiti/ [páasti] ‘to be tired/bored’

Suprasegmentals

The suprasegmentals in the Tucanoan languages are nasalization and tone or stress that is accompanied by high pitch.

Nasalization and Nasal Assimilation In all of the ET languages, nasalization is a feature of the morpheme. Root morphemes are either oral or nasal. Suffixes are either specified as oral or nasal, or they are unspecified for nasalization. Those that are unspecified are oral following an oral morpheme, and nasal following a nasal morpheme. In all of the Tucanoan languages, nasal spreading is progressive, spreading from left to right. In the following example from Des., /-re/ ‘specifier’ is unspecified for nasalization:

Des. (Miller, 1999: 14):

- (6) /igo-re/ [igore] ‘to her’
/bārī-re/ [mārīrē] ‘to us’

In the Middle Tucanoan language Cub., as in the ET languages, nasalization is a feature of the morpheme, but the nasal spreading rule is different. In Cub., nasality spreads through any suffix that begins with /b, d, j/. In Ret., nasal spreading is blocked by obstruents, and present analysis has indicated that it occurs within the metrical foot (Strom, personal communication).

In the WT languages, nasalization is a feature of the syllable and spreads through all suffixes that begin with /w, j, h, ʔ/. In addition, nasalization spreads through /dʒ/ in Kor. and through /hʷ/ in Sio. (Information is lacking on Ore.)

Regressive nasal spreading, where it occurs, is very limited. In the WT languages, it spreads from a morpheme consisting of a single nasal vowel to the preceding suffix. In the ET languages, regressive nasal spreading takes place in Des. (Miller, 1999: 14–15), Sir. (Criswell and Brandrup, 2000: 399), and Bas. (Jones and Jones, 1991: 15–16), affecting a limited set of specific, single-syllable suffixes. The regressive

spreading never goes beyond one syllable. In the Middle Tucanoan language Ret., regressive nasal spreading affects only the first-person singular morpheme, which is prefixed to the verb root (Strom, 1992: 20). To illustrate regressive nasal spreading, note what happens to the suffix /-bi/ 'negative' in the following examples:

Des. (Miller, 1999: 14):

(7) wēhē-bī-rā
kill-NEG-PL.ANIM
'the ones who don't kill'

(8) wēhē-bi-gi
kill-NEG-sing.MASC
'the one who doesn't kill'

Accent and Tone Of the 19 Tucanoan languages, 11 have tonal systems with high and low tone contrasting in identical or analogous environments. In Car., Tat., Mac., and Wai. (though perhaps not in Bar.), all four combinations of high and low on two syllable roots occur.

In Des., Tuy., and Yur., there is one accented syllable per phonological word, and it is associated with high pitch. Sec. and Sio. have a pattern of accent on alternate suffixes. Ret. has a system of multiple stress with rules that require epenthesis of suffixes and stress shifts (Strom, 1992: 13–19).

Accent is a property of the morpheme in the following ET languages:

- Bas. (Gómez-Imbert and Kenstowicz, 2000: 421)
- Des. (Miller, 1999: 15)
- Ret. (Strom, personal communication)
- Sir. (Criswell and Brandrup, 2000: 398)
- Tuc. (Ramirez, 1997: 68)
- Tuy. (Barnes, 1996: 31)
- Yur. (Kinch, personal communication)

The literature indicates that accent is probably a property of the morpheme also in:

- Cub. (Morse and Maxwell, 1999: 11–12)
- Mac. (Gómez-Imbert, 2000: 331)
- Pis. (González, 2000: 382).

Grammatical Characteristics

Sentence

The sentence in Tucanoan languages obligatorily demands a verb. Sentence fragments are used, but they occur, for example, as abbreviated answers to questions, responses using question words, etc.

Word Order In the majority of the Tucanoan languages, the usual word order in declarative clauses is

(S)(O)V, with variations according to discourse constraints. In Car., the preferred word order is (O)(S)V. Bas. exhibits the basic order OVS. Cub. also has (O)VS, but SV(O) occurs as frequently. Kor. is the only language that has a preferred word order in which the verb is initial: V(S)(O), although other word orders also occur. The pattern OV occurs in all the Tucanoan languages, and the languages exhibit other typical features of OV languages.

Case Markers Nouns in the role of grammatical subject are unmarked, and there are rules for when the complements are marked. In Ret., both an animate subject and an animate object may have the same marker /-re ~-te/. Where there could be confusion, it is avoided by word order: The subject precedes the object.

Ret. (Strom, 1992: 114):

(9) ernesto-te alvaro-te hedjobaa-rape
Ernest-HUM Alvaro-HUM help-PAST
'Ernest helped Alvaro'

Although the Tucanoan languages are almost exclusively suffixing languages, Car. and Tat. allow the complement, if it is a pronoun, to be prefixed to the verb, and Ret. has a neuter complement pronoun that only occurs as a prefix. In the rest of the Tucanoan languages, if the complement-pronoun needs to be expressed, it occurs in a separate word along with the complement/specificity suffix (REC = recent past; EV = evidential; NON3 = nonthird person – evidentiality and nonthird person are concepts discussed later in this article).

Car. (Gómez-Imbert, 2000: 332):

(10) ki-ĩjǎ-à-bō
3sing.MASC-see-REC-EV:PAST.VISUAL.3sing.FEM
'she saw him'

Sir. ('Brandrup, personal communication):

(11) sí-bī
give-EV:PAST.VISUAL.3sing.MASC
'he gave (it) (to me)'

(12) igó-re weré-bi
3sing.FEM-SPEC tell-EV:PAST.VISUAL.NON3
'I told (that) to her'

The specificity suffix, which marks significant participants and props in the discourse, does not always occur with nonpronominal direct objects. Noun incorporation is evident where the object precedes the verb root and is phonologically part of the verb word.

Des. (Miller, 1999: 109):

(13) diu-pi
egg-place.on.ground
'lay an egg'

Other nounverb combinations function as noun incorporation. Alternatively the noun may function as an independent complement (BEN = benefactive).

Tat. (Gómez-Imbert, 2000: 334):

- (14) *jì-pátu-kědōō- /kědōō-ja pátu-re*
 boha-ja
 1sing-coca-prepare- /prepare-IMP coca-SPEC
 BEN-IMP
 ‘prepare coca for me’ /‘prepare the coca’

When it is clear from the context that the noun is the complement, it is not marked as such, as in the case of ‘pigs’ in the following example (ANIM = animate, a concept discussed in this article).

Tuy. (Barnes, unpublished text #157):

- (15) *ape-‘bireko* ‘tuakūbū-adacu,
 other.day arrive.near.goal-FUT.1,2PL pig-PL
 je'se-a ja'a-adara hī't-ra
 eat-AFFIRM.1,2PL say-PL.ANIM
 ‘the next day we will arrive near (the town) in
 order to eat pigs’

With few exceptions, indirect objects, experiencers, and benefactors are always marked with the complement/specificity marker (SEP = separation, i.e., uncertain; D = dimension).

Mac. (Smothermon and Smothermon, 1995: 72):

- (16) *pauru-re iō-gi*
 Pablo-SPEC show-sing.MASC
 j-a jī
 AUXILIARY.VERB-PRES 1sing
 ‘I am showing (it) to Pablo’

Sio. (Wheeler, 2000: 187):

- (17) *jī'ŕt dī'hō-de gō'ŕā ha'ŕsi-gi-jā*
 1sing wife-SPEC bones hurt-certainty-SEP
 ‘my wife’s bones hurt’ (lit. ‘bones hurt my
 wife’)

Des. (Miller, 1999: 144):

- (18) *jī-re su'ŕi*
 1sing-SPEC clothes
 āsū-basa-ra-jē sājā-bi
 buy-BEN-NOM-CLASS: put.on-EV:PAST.VISUAL.NON3
 2D.flexible
 ‘I put on the dress (cloth) that was bought for me’

A separate set of suffixes identify location, time, instrument, and accompaniment (ACC).

Pis. (González, 2000: 387):

- (19) *wetŕe-pi*
 field-LOCATIVE
 ‘in the field’
 (20) *jābī-pi*
 night-LOCATIVE
 ‘at night’

- (21) *wābō-bēdā*
 hand-ISTR
 ‘by hand’
 (22) *kī-bai-bēdā*
 3sing.MASC-younger.brother-ACCOM
 ‘with his younger brother’

Nouns

Nouns in Tucanoan languages may be divided into two basic categories: animate and inanimate. These two categories take different plural suffixes. Within the animate category, human and nonhuman categories also take different plurals. Most ET nonhuman animate nouns are inherently singular and take a plural suffix. Nouns that refer to animals, insects, or fish that generally are found in groups are inherently plural and take a singularizing suffix.

Sir. (Criswell and Brandrup, 2000: 408):

- (23) *diá diá-rí*
 river river-PL.INAN
 ‘river’ ‘rivers’

Sir. (Criswell and Brandrup, 2000: 405):

- (24) *bāhí-gī bāhí-rā*
 child-sing.MASC child-PL.ANIM
 ‘boy’ ‘children’
 (25) *pābú pābú-ā*
 armadillo armadillo-PL.ANIM
 ‘armadillo’ ‘armadillos’
 (26) *burúa burúa-bī*
 termites termites-SINGULARIZER
 ‘termites’ ‘a termite’

WT animate nouns have three categories: general, singular (with a singular suffix that is either masculine or feminine) and plural (with a plural suffix).

Sio. (Wheeler, 2000: 185):

- (27) *'zī 'zīgī 'zīgo 'zīk"ā*
 ‘child’ ‘boy’ ‘girl’ ‘children’

Classifiers The Tucanoan languages all have a small set of animate classifiers and, in most of the languages, a larger set of inanimate classifiers.

Animate classifiers, which also function as animate nominalizers, include masculine singular, feminine singular, and plural. Most of the languages have past, present, and future forms for the animate nominalizers as is shown in Table 4 for Sir.

Sir. (Criswell and Brandrup, 2000: 408):

- (28) *bueé-gi*
 study-sing.MASC
 ‘he who studies’

Table 4 Animate classifiers in Siriano (Criswell and Brandrup, 2000: 408)

| <i>Tense</i> | <i>Singular, Masculine</i> | <i>Singular, Feminine</i> | <i>Plural</i> |
|--------------|----------------------------|---------------------------|---------------|
| Present | -gi | -go | -rã |
| Past | -dii-gi | -dee-go | -dēē-rã |
| Future | -bu-gi | -bu-go | -bū-rã |

- (29) buué-dii-gi
study-PERFECTIVE-sing.MASC
 'he who studied'

- (30) buué-bu-gi
study-POTENTIAL-sing.MASC
 'he who will study'

Typically each ET language has over 100 inanimate classifiers, plus many nouns that may also function as classifiers. Ramirez (1997: 109) says that Tuc. has only six classifier suffixes, which he calls "shape suffixes" (plus some 400 'dependent nouns'). The 'shape suffixes' that both West and Welch (2000: 428 and personal communication) and Ramirez have identified are a small set of classifiers that do not require a nominalizer between the verb root and the 'shape suffix,' as do the many Tuc. suffixes that correspond to classifiers found in other ET languages.

Inanimate classifiers in Tucanoan languages typically are suffixes that categorize the object(s) referred to in terms of some salient characteristic, such as shape or arrangement. For a complete description of Tuy. classifiers see Barnes (1990). Tucanoan classifiers have been variously described as noun classifiers or numeral classifiers. In the ET languages, they are suffixed to numerals, demonstratives, quantifiers, genitives, nouns, and to either nominalized verbs and/or descriptive adjectives, or, in the case of the WT languages, directly to these roots as nominalizers.

Des. (Miller, 1999: 37, 5, 45, 125, 38, 40):

- (31) juhu-koaru
one-CLASS:gourd
 'one gourd'
- (32) iri-ru
this-CLASS:oblong
 'this airplane'
- (33) baha-bihĩ-ri
a.lot-CLASS:thin.plane-PL.INAN
 'many knives'
- (34) bĩ-ja-ru
2sing-GEN-CLASS:oblong
 'your boat'
- (35) juki-kawe
tree-CLASS:bent
 'crooked tree'

- (36) oʔa-ri-boga
sweep-NOM-CLASS:bundle
 'broom'

When classifiers are suffixed to nouns, they are in a relationship of General-specific, as in the following examples, where /dēĩ/ is the 'mirití' palm.

Cub. (Ferguson *et al.*, 2000: 361):

- (37) dēĩ-jĩ 'palm tree'
 dēĩ-ri 'palm fruit'
 dēĩ-kũ 'cluster of palm fruit'
 dēĩ-jabe 'seed of the palm fruit'
 dēĩ-joka 'palm leaf'

The WT languages have from 17 (Sec.) to 30 (Kor.) classifiers. These classifiers are suffixed to nouns, as in the Cub. examples above, and to verbs and adjectives as nominalizers. They are sometimes found suffixed to numerals and demonstrative adjectives.

The more than 29 Ret. classifiers are suffixed to nouns, numerals, nominalized verbs and descriptive adjectives, and optionally to demonstrative adjectives.

Noun Modifiers Noun modifiers in Tucanoan languages may be divided into two major groups: limiting adjectives and nominalized adjectival verbs.

Limiting adjectives, which are numerals, genitives, demonstratives (anaphoric or exophoric), and quantifiers, are either separate words or roots that require suffixes.

Mac. (Smothermon and Smothermon, 1995: 39–40):

- (38) hĩa-hibi
two-CLASS:basket
 'two baskets'
- (39) hĩgi ĩ ja-gĩ
hammock 3sing.MASC GEN-CLASS:hammock
 'his hammock'

Yur. (Kinch and Kinch, 2000: 477):

- (40) ai-wi
that(EXOPHORIC)-CLASS:building
 'that house'

Des. (Miller, 1999: 45):

- (41) baha-bē-rã
a.lot-NEG-PL.ANIM
 'a few (people)'

Nominalized adjectival verbs take the place of descriptive adjectives in the traditional sense of the term. Stative verbs, such as 'to be red' and 'to be big,' do not always take the full range of verb suffixes, and yet they are used as verbs, as can be seen in the following example.

Tuy. (Barnes, unpublished text #82):

- (42) 'jaa dia'poa baji'ro
 1_{GEN} face very
 ji'i-a
 be.black-EV:PRES.VISUAL.NON3
 'my face is really dark (from the sun)'

Nominalized adjectival verbs may serve as full constituents of the sentence, i.e., the noun that is being modified will not appear in the sentence if the referent is already clear. When the referent is an animate noun, an animate classifier is suffixed directly to the verb and functions as a nominalizer. In the case of an inanimate referent, some of the languages require that a nominalizer be suffixed to the verb before the classifier.

Wan. (Waltz and Waltz, 2000: 460):

- (43) já-ĩdā
 be.bad-CLASS:PL.ANIM
 'bad people'

Bas. (Jones and Jones, 1991: 63):

- (44) sũā-ri-hāĩ joa-ri-hāĩ
 be.red-NOM-CLASS:2D be.long-NOM-CLASS:2D
 ābō-a-ha jĩ
 want-PRES-NON3 1sing
 'I want a long red piece of cloth'

Cub. has a small class of descriptive adjectives, which function neither as nouns nor as verbs. Among these are: *big*, *small*, *old*, *dry*, and *curly* (Morse and Maxwell, 1999: 124). Des., Sir., and Tuc. each list a small number of adjectives, among which are: *big* and *small* (Miller, 1999: 51; Criswell and Brandrup, 2000: 410; Welch, personal communication). Sec. and Sio. also each have a small class of descriptive adjective roots, including *big* and *small*, and derive the rest of their descriptive words from other grammatical forms (Johnson and Levinsohn, 1990: 37–38; Wheeler, 1987: 116–117). In Ret., words that are traditionally thought of as descriptive adjectives function almost exactly like nouns, and thus are listed as nouns (Strom, 1992: 23–26). The rest of the Tucanoan languages derive descriptive words by nominalizing verb roots and/or suffixing gender/number or classifier suffixes. In many, but not all, cases, these function much as do relative clauses in other languages.

Cub. (Morse and Maxwell, 1999: 86):

- (45) xidoxa-RI-xārāwĩ
 be.scary-NOM-CLASS:day
 'a scary day'

Personal Pronouns All of the Tucanoan languages have the same system for personal pronouns. In the singular, there are four forms: first person, second

person, third-person masculine and third-person feminine. In the plural, there are also four forms: first-person exclusive, first-person inclusive, second person, and third person.

Tat. exhibits a typical set as shown in Table 5.

Demonstrative Adjectives The Tucanoan languages are split as to whether they make a distinction between singular and plural in the demonstrative adjectives for inanimate referents. Cub., for example, does not make the distinction: /i-/ means both 'this' and 'these'; /ādĩ-/ means both 'that' and 'those.' The anaphoric pronoun /di-/ means both 'that' and 'those' (Morse and Maxwell, 1999: 96). Pluralization is indicated only on the classifier or noun that follows the demonstrative adjective. Tuy. does make the distinction as shown in Table 6, and thus number is indicated on both the demonstrative adjective and the classifier or noun which follows it.

Tuy. (my field data):

- (46) ati-do'to
 'this-CLASS:large.bundle
 'this large bundle (of firewood, cane, etc.)'
 (47) ate-do'to-ri
 this.PL-CLASS:large.bundle-PL.INAN
 'these large bundles (of firewood, cane, etc.)'

Verbs

Independent verbs in Tucanoan languages are minimally comprised of a verb root and an evidential

Table 5 Personal pronouns in Tatuyo (Gómez-Imbert, 2000: 341)

| | Person | Singular | Plural |
|---|-----------|--------------------------|-----------------------------|
| 1 | exclusive | | <i>hāā</i> |
| | inclusive | <i>jii</i> | |
| 2 | masculine | <i>bĩĩ</i> <i>kĩĩ</i> | <i>bāđĩ</i> <i>bĩhāā</i> |
| | feminine | <i>kóō</i> | <i>dāā</i> |

Table 6 Demonstrative pronouns in Tuyuca (Barnes and Malone, 2000: 446)

| Pronoun | Singular | Plural/Noncountable |
|-----------|-------------|---------------------|
| Exophoric | | |
| 'this' | <i>ati-</i> | <i>ate</i> |
| 'that' | <i>ii-</i> | <i>iye</i> |
| Anaphoric | | |
| 'that' | <i>tii-</i> | <i>tee</i> |

suffix, or an imperative, interrogative, or future suffix. In Tat. and Car., the verb root may be prefixed by a pronoun, and, to a lesser extent, the same is true for Bas. and Wai. Mood is indicated by suffixes that occur between the verb root and the evidential suffix, and may include negative, contraexpectation, desiderative, and irrealis suffixes. Some aspects are also indicated by verb suffixes that occur between the verb root and the evidential suffix, and indicate, for example, the habitual, durative, completive, and iterative aspects. Other aspects are indicated by auxiliary verb phrases.

Auxiliary Verbs The most common use of an auxiliary verb is in expressions of the progressive aspect. The main verb is nominalized, and the auxiliary verb is suffixed by dependent or independent verb suffixes, including whatever aspect or mood suffixes may be appropriate to the situation.

Tuy. (my field data):

- (48) 'waa-gi tii-'bī-wi
 go-NOM: do-CONTRAEXPECTATION-
 sing.MASC EV:PAST.VISUAL.sing.MASC
 'He was going, but ...'

Compound Verb Roots Compounding of verb roots is rare in the WT languages and Cub., but common in Ret. and the ET languages, where up to four verb roots can be combined in one phonological word. See Gómez-Imbert (1988) for an explanation of three different types of compounding that typically take place.

Tat. (Gómez-Imbert, 1988: 107):

- (49) yáá-róka-kûbû-ehá
 fall-strike-lie.immobile-arrive
 'to fall, arriving at and striking (the ground),
 and being immobile'

Evidentiality Evidentiality is an obligatory feature of the independent verb word in the ET languages, as well as in Cub., Sio., and Sec. It is indicated by optional verb suffixes in Ret., and is indicated by means of auxiliary verbs in Kor.

Evidential suffixes in the ET languages carry information about present and past tenses, and subject, including person, number, and gender. The function of the evidential suffixes appears to vary somewhat between the languages, indicating one of the following:

1. How the speaker obtained the information: Bar. and Wai. (my field data); Bas., Car., Mac., and Tat.

(Gómez-Imbert, 2000: 340); Des. (Miller, 1999: 64); Sir. (Criswell and Brandrup, 2000: 400); Tuy. (Barnes and Malone, 2000: 441); Yur. (Kinch and Kinch, 2000: 479),

2. The speaker's degree of knowledge of the situation: Tuc., according to Ramirez (1997: 121), or
3. the point of view of the speaker: Wan. (Waltz and Waltz, 2000: 456); Tuc., according to Welch and West (2000: 424).

Evidential suffixes in the Middle Tucanoan language Cub. and in the WT languages Sio. and Sec. convey tense and person information as in the ET languages, but they indicate the degree of certainty about the information rather than how the speaker obtained the information (Ferguson *et al.*, 2000: 363; Wheeler, 2000: 189; Johnson and Levinsohn, 1990: 66–70). There is no information available on Ore.

Ret.'s evidential system consists of three optional verb suffixes. The first is /-ko-/, by which the speaker tells something that he knows is a fact because he has heard something take place, although he has not seen it. The second is /-rihi-/, by which the speaker indicates that he is stating an assumption. The third is /-re/, by which the speaker conveys that the information is secondhand (Strom, 1992: 90–91).

Kor. employs auxiliary verbs to indicate evidentiality. One indicates secondhand information and the other indicates a supposition on the part of the speaker (Cook and Criswell, 1993: 86–87).

Wan. (Waltz and Waltz, 2000: 457) and Des. (Miller, 1999: 67–68) use an auxiliary verb phrase for the 'apparent' evidential, which is equivalent in function to the apparent evidential suffix in Tuy. The cognate verb phrase in Tuy. is distinct from the apparent evidential; the auxiliary verb bears the witnessed evidential suffix, and indicates that the speaker visually observed the end result of an action or state. (See Barnes, 1984: 264). Thus, by looking into the empty house he says, 'They left,' literally saying, 'I see that they are ones who have left.' Tuy. also has a single-syllable evidential suffix that indicates 'apparent' actions or states that the speaker deduces from evidence. That single-syllable suffix would be used, for example, when concluding that a piece of fruit was **apparently** in the state of being ripe, or it would not have fallen off the tree on its own. In the text where the following example occurs, the speaker heard the fruit fall, but never saw it.

Tuy. (Barnes, unpublished text #137):

- (50) yīi-'ri-ga dīi-bī-a-ju
 ripen-NOM- be-CONTRAEXPECTATION-REC-
 CLASS:3D EV:PAST APPARENT.NON3
 'apparently the fruit was ripe'

Table 7 Evidentials in Siriano (Criswell and Brandrup, 2000: 400)

| <i>Tense</i> | | <i>Visual</i> | <i>Apparent</i> | <i>Secondhand</i> | <i>Assumed</i> |
|--------------|------------|---------------|-----------------|-------------------|-----------------|
| Past | NON3 | - <i>bī</i> | - <i>jo</i> | - <i>juro</i> | - <i>kujo</i> |
| | 3sing.MASC | - <i>bī</i> | - <i>jūbī</i> | - <i>jupi</i> | - <i>kūjūbī</i> |
| | 3sing.FEM | - <i>bō</i> | - <i>jūbō</i> | - <i>jupo</i> | - <i>kūjūbō</i> |
| | 3PL | - <i>bā</i> | - <i>jūbā</i> | - <i>jūrā</i> | - <i>kūjūbā</i> |
| Present | NON3 | - <i>a</i> | - | - | - <i>koa</i> |
| | 3sing.MASC | - <i>bī</i> | - | - | - <i>kūbī</i> |
| | 3sing.FEM | - <i>bō</i> | - | - | - <i>kūbō</i> |
| | 3PL | - <i>bā</i> | - | - | - <i>kūbā</i> |

Malone has concluded that the apparent evidential in Tuy. developed as the Tuy. speakers moved from expressing speaker distance in time and space to emphasizing how the speaker obtained his information (Malone, 1988: 138).

Table 7 illustrates a typical Tucanoan evidential system.

As is typical of most of the ET languages, person markers distinguish between third person and non-third persons. In Table 7, NON3 includes first and second persons singular and plural, plus inanimate.

Note that two of the eight paradigms in Table 7 are totally incomplete. The present tense in Sir. only distinguishes between the visual and assumed evidentials. The largest and most complete set of paradigms described in the literature on ET languages is found in Tuy., where 8 of 10 paradigms are complete (Barnes, 1984: 258). One of the incomplete paradigms is missing only the NON3 suffix. The totally empty paradigm is the present secondhand slot where one would not expect a paradigm, since secondhand information is always reported as past. Note the use of present and past in the following examples:

Tuy. (my field data):

(51) Bar'ia a'ti-jo

Maria come-EV: PRES.VISUAL.3sing.FEM

'Maria is coming' (reported by a person outside the house who sees Maria coming)

(52) a'ti-a-jigo

come-REC-EV: PAST.SECONDHAND.3sing.FEM

'she is coming' (lit. 'I was told that she was coming'), (reported by a person inside the house who has not seen Maria, but who heard the other say that she is coming)

Two features of evidentiality that occur in all of the Tucanoan languages are (1) firsthand knowledge of the state or event and (2) secondhand knowledge, indicating that the only information the speaker has

Table 8 Probable and indefinite future suffixes in Yuruti (Kinch and Kinch, 2000: 480, and personal communication)

| <i>Person</i> | <i>Probable future</i> | <i>Indefinite future</i> |
|---------------|------------------------|------------------------------------|
| 1,2 sing.MASC | - <i>giaku</i> | - <i>giga</i> |
| 1,2 sing.FEM | - <i>goaku</i> | - <i>goga</i> |
| 1,2PL | - <i>roaku</i> | - <i>raga</i> |
| 3sing.MASC | - <i>giaki</i> | - <i>giagawi</i> |
| 3sing.FEM | - <i>goakugo</i> | - <i>goagago</i> |
| 3PL | - <i>rakua</i> | - <i>ragawa</i> ~ - <i>roagawa</i> |
| Inanimate | - <i>roaku</i> | - <i>roga</i> |

about what he relates is that which came from someone else. All other features such as direct or indirect evidence, and tangible or intangible evidence, or, in the WT languages: degree of certainty, can be subdivisions of point (1).

Future The WT languages express the future by means of the potential aspect. The Middle and ET languages express the future in modal terms. For example, Bas. uses one of three moods to indicate the future: avoidance, conjecture, and intention (Jones and Jones, 1991: 88–92). The rest of the Tucanoan languages have from one to three sets of future endings, composed of two to three morphemes each, which together convey probability, supposition, or definite intention. A study of Table 8 reveals how more than one morpheme is typically used in the formation of the future tenses.

Yur. (Kinch and Kinch, 2000: 480):

(53) bē'dābē-pi 'waa-giaki
tomorrow-LOCATIVE go-FUT.PROBABLE
 'he will go tomorrow'

(54) ati-jā'bika jo'sa-giagawi 'kꞑ
this- hang.in.hammock- 3sing.MASC
afternoon FUT.INDEFINITE
 'probably he will rest in his hammock this afternoon'

Bibliography

- Aikhenvald A (2002). *Language contact in Amazonia*. New York: Oxford University Press.
- Barnes J (1970–2005). 'Unpublished texts.'
- Barnes J (1984). 'Evidentials in the Tuyuca verb.' *International Journal of American Linguistics* 50(3), 255–271.
- Barnes J (1990). 'Classifiers in Tuyuca.' In Payne D L (ed.) *Amazonian linguistics, studies in Lowland South American Languages*. Austin: University of Texas Press. 273–292.
- Barnes J (1996). 'Autosegments with three-way lexical contrasts in Tuyuca.' *International Journal of American Linguistics* 62(1), 31–58.
- Barnes J & Malone T (2000). 'El tuyuca.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 437–452.
- Cook D M & Criswell L L (1993). *El idioma koreguaje (Tucano Occidental)*. Santafé de Bogotá: Asociación Instituto Lingüístico de Verano.
- Criswell L & Brandrup B (2000). 'Un bosquejo fonológico y gramatical del siriano.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 395–417.
- Ferguson J, Hollinger C & Criswell L (2000). 'El cubeo.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 357–372.
- Gómez-Imbert E (1988). 'Construcción verbal en barasana y tatuyo.' *Amerindia* 13, 97–108.
- Gómez-Imbert E (2000). 'Introducción al estudio de las lenguas del Piraparaná (Vaupés).' In González de Pérez M S & Rodríguez de Montes M L (eds.). 321–356.
- Gómez-Imbert E & Kenstowicz M (2000). 'Barasana tone and accent.' *International Journal of American Linguistics* 66(4), 419–463.
- González de Pérez M S (2000). 'Bases para el estudio de la lengua pisamira.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 373–393.
- González de Pérez M S & Rodríguez de Montes M L (eds.) (2000). *Lenguas indígenas de Colombia, una visión descriptiva*. Santafé de Bogotá: Instituto Caro y Cuervo.
- Johnson O E & Levinsohn S H (1990). *Gramática secoya*. Quito: Instituto Lingüístico de Verano.
- Jones W & Jones P (1991). *Studies in the languages of Colombia 2: Barasano syntax*. Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.
- Kinch R & Kinch P (2000). 'El yurutí.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 469–487.
- Klumpp J & Klumpp D (1973). 'Sistema fonológico del piratapuyo.' In *Sistemas fonológicos de idiomas colombianos*, tomo 2. Colombia: Instituto Lingüístico de Verano. 107–120.
- Malone T (1986). *Proto-Tucanoan and Tucanoan genetic relationships*. (MS).
- Malone T (1988). 'The origin and development of Tuyuca evidentials.' *International Journal of American Linguistics* 54(2), 119–140.
- Metzger R G (1998). 'The morpheme KA- of Carapana (Tucanoan).' *SIL Electronic Working Papers* [on line], April, Available: <http://www.sil.org/>.
- Miller M (1999). *Studies in the languages of Colombia 6: Desano grammar*. Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.
- Morse N L & Maxwell M B (1999). *Studies in the languages of Colombia 5: Cubeo grammar*. Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.
- Nagler C & Brandrup B (1979). 'Fonología del siriano.' In *Sistemas fonológicos de idiomas colombianos 4*. Colombia: Instituto Lingüístico de Verano. 101–126.
- Ramírez H (1997). *A fala tukano dos ye'pa-masa. Tomo I: Gramática*. Manaus: CEDEM.
- Smothermon J R, Smothermon J H & con Frank P S (1995). *Bosquejo del macuna*. Santafé de Bogotá: Asociación Instituto Lingüístico de Verano.
- Stenzel K (2004). *A reference grammar of Wanano*. Ph.D. diss., University of Colorado.
- Strom C (1992). *Studies in the languages of Colombia 3: Retuarã syntax*. Dallas: The Summer Institute of Linguistics and The University of Texas at Arlington.
- Waltz N E (2002). 'Innovations in Wanano (Eastern Tucanoan) when compared to Piratapuyo.' *International Journal of American Linguistics* 68(2), 157–215.
- Waltz C & Waltz N (2000). 'El wanano.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 453–467.
- Waltz N E & Wheeler A (1972). 'Proto Tucanoan.' In Matteson E *et al.* (eds.) *Janua Linguarum 127: Comparative studies in Amerindian languages*. The Hague: Mouton. 119–149.
- Welch B & West B (2000). 'El tucano.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 419–436.
- West B & Welch B (1967). 'Phonemic system of Tucano.' In Elson B F (ed.) *Phonemic systems of Colombian languages*. University of Oklahoma: Summer Institute of Linguistics. 11–24.
- Wheeler A (1987). *Ganteya bain, el pueblo siona del río Putumayo, Colombia*, tomo 1. Colombia: Instituto Lingüístico de Verano.
- Wheeler A (2000). 'La lengua siona.' In González de Pérez M S & Rodríguez de Montes M L (eds.). 181–198.

Tungusic Languages

A Vovin, University of Hawaii at Manoa, Honolulu, HI, USA

© 2006 Elsevier Ltd. All rights reserved.

Location and Composition of the Tungusic Language Family: Sociolinguistic Data

Tungusic languages (former name: Manchu–Tungusic) are spoken in a wide territory that includes Central and East Siberia and northeast China (Manchuria). One language, Sibe (Xibe), is located in the Xinjiang province of China. There are 12 modern Tungusic languages (see Table 1).

There are also two languages attested diachronically: Jurchen, represented by some inscriptions and dictionaries (12th–15th centuries C.E.), and Manchu, the state language of the Qing empire in China (1644–1911 C.E.) that has an abundant corpus of various written texts (16th to the early 20th century C.E.), most of them translations from Chinese. For all practical purposes, Manchu can be considered an archaic dialect of Jurchen, although the two languages employ different writing systems. Jurchen used a cumbersome indigenous (inspired by Chinese and Khitan writing systems) script that included both semantographic and syllabic signs. Manchu, on the other hand, uses the modified version of the Mongolian alphabet. Classical Manchu is still used in private correspondence by Heilongjiang Manchu, Solon, and Dagur Mongolians.

All surviving Tungusic languages are endangered to a greater or lesser extent; some of them are on the brink of extinction. The only languages spoken in China that have a written form are (1) Heilongjiang Manchu, whose speakers use essentially Classical Manchu, and (2) Sibe, whose speakers use a modified

version of Classical Manchu that includes certain Sibe colloquialisms, because Sibe is historically a dialect of Manchu. During the Soviet period, literary forms were created for almost all Tungusic languages spoken in Russia, but most of them turned out to be short-lived. Even those languages that still have literary forms (Ewenki, Ewen, and Nanai) have a rather narrow application.

Internal Classification

There have been several conflicting attempts to classify Tungusic languages, but the most convincing is the recent attempt by Stefan Georg (2004), in which he proposed two basic groups: Northern (Ewenki, Ewen, Solon, Neghidal, Udehe, and Oroch) and Southern (Nanai, Ulcha, Uilta, both Classical and Modern Manchu, Sibe, and Jurchen). Within the Northern group, the following three subgroups should be distinguished: Ewen, Ewenki–Solon–Neghidal, and the intermediate subgroup including Udehe and Oroch. The languages of the intermediate subgroup are basically the Northern Tungusic languages that were strongly influenced by the Southern Tungusic languages. The Kili language that was traditionally considered a dialect of Nanai also likely belongs to the Northern group (Doerfer, 1978a). There are two subgroups within the Southern group: the Nani subgroup, including Nanai, Ulcha, and Uilta; and the Manchu subgroup, which included Manchu (both Classical and Modern), Sibe, and Jurchen.

Wider Genetic Affiliation

There is a widespread belief that Tungusic languages are distantly related to Mongolic, Turkic,

Table 1 Modern Tungusic languages

| | <i>Names appearing in Ethnologue</i> | <i>Frequently used alternative names</i> | <i>Number of native speakers^a</i> |
|---------------------|--------------------------------------|--|--|
| Ewenki | Evenki | Elunchun | 11 360 |
| Ewen | Even | Lamut | 7463 |
| Solon | Evenki | Ewenke | 17 000 |
| Neghidal | Negidal | | 170 |
| Kili | Kili | Kur-Urmi, Hezhen | 40 |
| Udehe | | Udige | 526 |
| Oroch | Oroch | | 169 |
| Nanai | | Goldi | 5292 |
| Ulcha | Ulcha | Olcha | 986 |
| Uilta | | Orok | 89 |
| Heilongjiang Manchu | Manchu | | 70 |
| Sibe | Xibe | Xibo | 26 760 |

^aEstimates of native speakers based on Soviet census of 1989 and Chinese census of 1990.

Korean, and Japonic languages, forming with them the Altaic family. However, this controversial relationship has never been demonstrated satisfactorily. It is most likely that numerous parallels between the Tungusic and other Altaic languages represent traces of centuries- or even millennia-long contacts.

Structure

All Tungusic languages are agglutinative (with some elements of fusion) languages with SOV word order, although Ewen in some cases has shifts to SVO order, apparently under a strong Russian influence. Thus, there is only suffixation and no prefixation. Almost all languages have a rich morphology, with a somewhat reduced version of it in the Manchu subgroup.

Phonology

Tables 2 and 3 show the vowels and consonants for the Podkamennaia Tunguska subdialect of the Southern Ewenki dialect, which is used as the basis of the modern literary language. Both vocalic and consonantal systems are representative of the whole family, although, of course, certain expansions and/or reductions can be observed in individual languages. Syllabic structure is V, VC, CV, and CVC. Stress is probably dynamic, although further research is necessary. All languages have vowel harmony.

All vowels can be either short or long except *e* (< diphthong **ia*), which is always long. Vowel length is phonemic (cf. Ewenki *bu-* ‘to die,’ *bū-* ‘to give’; *tūkala* ‘name of a plant,’ *tukala* ‘earth, ground’).

Table 2 Vowels in Tungusic languages

| | Front | Central | Back |
|------|-------|---------|--------|
| High | i, ī | | u, ūa |
| Mid | ē | ə, ǣ | o, ō |
| Low | a, ā | | |

Table 3 Consonants in Tungusic languages

| | Bilabial | | Dental | | Palatal | | Velar | | Glottal |
|---------------------|----------|---|--------|---|---------|---|-------|---|---------|
| Plosive | p | b | t | d | c | ʃ | k | g | |
| Nasal | | m | | n | | ɲ | | ŋ | |
| Trill | | | | r | | | | | |
| Fricative | | β | s | | | | | | h |
| Approximant | | | | | j | | | | |
| Lateral approximant | | | | l | | | | | |

Morphology

Overall, the Northern Tungusic languages have a richer morphology than the Southern Tungusic languages. Nouns in most Tungusic languages have categories of number, case, and possession, although possession is not present in the Manchu subgroup. The number of cases varies from 6 in Manchu to 13 in Ewen. There is a certain allomorphy in case suffixes, depending on the last consonant of a nominal stem and vowel harmony. Table 4 shows an Ewenki paradigm that includes 11 cases for the words *bira* ‘river,’ *dət* ‘tundra,’ and *oron* ‘reindeer.’

Some Tungusic languages differentiate between alienable and inalienable possession (cf. Ewenki *dili-B* head-1PERS.sing.POSS ‘my head’ and *dili-ŋi-β* head-ALIEN-1PERS.sing.POSS ‘head of an animal that I killed and have in my possession,’ where alienable possession is indicated by the special affix -*ŋi-*). There is also a distinction between exclusive and inclusive first person plural pronouns (cf. Manchu *be* ‘we without you’ and *muse* ‘we including you,’ ‘I and you’).

In some languages, adjectives agree with the modified noun in number and case, as for example in Ewenki: *ərū-l-dū bira-l-dū* bad-PL-DAT.LOC river-PL-DAT.LOC ‘in bad rivers’; adjectives stay uninflected in other languages, as in Nanai: *dāi xoton-sal-jiafi* big city-PL-EL ‘from big cities.’

The verbal morphology is very complex. All languages differentiate between nonfinite and finite verbal forms. Verbs have the following categories: voice, aspect, mood, tense, person, and number. There are six moods, six voices, and ten different aspects in Ewenki. The typical order of affixes in a verbal form is root-VOICE-ASPECT-MOOD-TENSE-PERSON/NUMBER (e.g., Ewenki *ana-wkān-ʃə-cə-n* push-CAUS-IMPERF-PAST-3PERS.sing ‘she was making [him] to push’). In most Tungusic languages, there is a special negative verb (e.g., Ewenki *baka-ra-n* find-AOR.PART-3PERS.sing ‘he found,’ *ə-cə-n baka-ra* NEG.V-PAST-3PERS.sing find-AOR.PART ‘he did not find,’ *baka-ʃaŋā-n* find-FUT-3PERS.sing ‘he will find,’ *ə-ʃaŋā-n baka-ra* NEG.V-FUT-3PERS.sing find-AOR.PART ‘he will not find’).

Table 4 Morphology in Tungusic languages

| | Vowel stem | Plosive stem | Nasal stem |
|-----------------------|------------|--------------|------------|
| Nominative | bira | dət | oron |
| Accusative | bira-βa | dət-pe | oron-mo |
| Indefinite accusative | bira-ja | dət-je | oron-o |
| Dative–locative | bira-dūa | dət-tūa | oron-dūa |
| Allative | bira-tki | dət-tiki | oron-tiki |
| Illative | bira-lā | dət-[tu]lā | oron-dulā |
| Prolative | bira-lī | dət-[tu]lī | oron-dulī |
| Allative–locative | bira-kla | dət-iklā | oron-ikla |
| Elative | bira-duk | dət-tuk | oron-duk |
| Ablative | bira-git | dət-kit | oron-ŋit |
| Instrumental | bira-t | dət-it | oron-di |

Bibliography

- Alpatov V M, Kormushin I V, Piurbuev G Ts & Romanova O I (eds.) (1997). *Iazyki mira: Mongol'skie iazyki, tunguso-man'chzhurskie iazyki, iaponskii iazyk, koreiskii iazyk*. Moscow: Indrik.
- Benzing J (1955). *Die Tungusischen Sprachen: Versuch einer vergleichender Grammatik*. Mainz: Verlag der Akademie der Wissenschaften und der Literatur.
- Chao Ke D O (1997). *Man – tonggusi zhe yu: pijiao yanjiu*. Beijing: Minzu chubanshe.
- Doerfer G (1978a). 'Classification problem of Tungus.' In Doerfer G & Weiers M (eds.) *Tungusica, Band 1: Beiträge zur Nordasiatischen Kulturgeschichte*. Wiesbaden: Otto Harrassowitz. 1–26.
- Doerfer G (1978b). 'Urtungusisch *ö.' In Doerfer G & Weiers M (eds.) *Tungusica, Band 1: Beiträge zur Nordasiatischen Kulturgeschichte*. Wiesbaden: Otto Harrassowitz. 66–116.
- Doerfer G (1985). *Mongolo–Tungusica (Tungusica, Band 3)*. Wiesbaden: Otto Harrassowitz.
- Georg S (2004). 'Unreclassifying Tungusic.' In Naeher C (ed.) *Proceedings of the First International Conference on Manchu-Tungus Studies. 2: Trends in Tungusic and Siberian linguistics (Tunguso-Sibirica 9)*. Wiesbaden: Otto Harrassowitz. 45–57.
- Ikegami J (1989). 'Tsongûsu shogo.' *Gengogaku daijiten II*, 1058–1083.
- Ikegami J (2001). *Tsongûsu go kenkyû*. Tokyo: Kumifuru shoin.
- Kazama Sh (2003). *Basic vocabulary (A) of Tungusic languages*. Osaka: Endangered Languages of the Pacific Rim (ELPR).
- Skorik P I, Avrorin V A, Bertagaev T A, Menovshchikov G A, Sunik O P & Konstantnova O A (eds.) (1968). *Iazyki narodov SSSR, 5: Mongol'skie, tunguso-man'chzhurskie i paleoaziatskie iazyki*. Leningrad: Nauka.
- Sunik O P (1962). *Glagol v tunguso-man'chzhurskix iazykakh*. Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Sunik O P (1982). *Sushchestvitel'noe v tunguso-man'chzhurskix iazykakh*. Leningrad: Nauka.
- Tsintsius V I (1949). *Sravnitel'naia fonetika tunguso-man'chzhurskikh iazykov*. Leningrad: Uchpedgiz.
- Tsintsius V I (ed.) (1975–1977). *Sravnitel'nyi slovar' tunguso-man'chzhurskikh iazykov*, (vols 1 & 2). Leningrad: Nauka.
- Tsumagari T (1983). 'Tsongûsu go.' *Gekkan gengo* 12(11).
- Vovin A (1993). 'Towards a new classification of Tungusic languages.' *Eurasian Yearbook* 65, 99–113.
- Vovin A (ed.) (in press). *The Tungusic languages*. Routledge Language Family Series. London: Routledge.

Tupian Languages

N Gabas Jr, Belém, Brazil

© 2006 Elsevier Ltd. All rights reserved.

The Tupí family is one of the largest families of languages of South America. It contains 10 branches, with a variety of languages in each branch. The first comprehensive classification of the Tupian languages was by Rodrigues (1964), and further improvements of his classification were made by Cabral (1996, 1997), Gabas (2000), Rodrigues and Cabral (2002), Rodrigues and Dietrich (1997), and Rodrigues (1966, 1980, 1985a, 1997). It is generally accepted that the point of origin of Tupian groups is the state of Rondônia, in the northwest part of Brazil. Rondônia is still the homeland of five Tupian branches – Arikém, Mondé, Puruborá, Ramaráma,

and Tuparí – and of a few dialects (Amondawa, Karipuna, and Uru-eu-wau-wau) of the Kawahib cluster of the Tupí-Guaraní branch.

Nine branches of the Tupí family are shown in Table 1, together with the languages that belong to each branch. Classification of the tenth branch of the Tupí family, Tupí-Guaraní, is shown separately, in Table 2, because of its complexity; the Tupí-Guaraní branch has the largest number of languages of the Tupí family (almost 50 languages, arranged in several subgroupings), and several of its members are spoken in countries other than Brazil. In Table 1, languages on the same line separated by a slash correspond to dialects of the same language; languages within parentheses correspond to alternate names for that language. In Table 2, language clusters are indicated by italics. These correspond roughly to dialects of the same language. The population numbers given in both

Table 1 Classification of nine branches of the Tupi family

| <i>Branch</i> | <i>Language</i> | <i>Population</i> |
|---------------|----------------------------------|------------------------|
| Awetí | Awetí | 100 |
| Arikém | Arikém | Extinct |
| | Karitiána | 170 |
| Jurúna | Jurúna | 210 |
| | Xipáya | 15 (two speakers) |
| Mawé | Mawé (or Sateré) | 5800 |
| Mondé | Aruá/Cinta-Larga/Gavião/ Zoró | 36/640/360/250 |
| | Mondé (Salamã) | 3 (semi-speakers) |
| | Suruí | 580 |
| Mundurukú | Mundurukú | 3000 |
| | Kuruáya | 10 |
| Puruborá | Puruborá | 20 (two semi-speakers) |
| Ramaráma | Karo (Arara) | 170 |
| Tuparí | Ayurú | 40 |
| | Akuntsu | 7 |
| | Makuráp | 130 |
| | Mekéns (Sakirabiat) | 70 |
| | Tuparí | 200 |

tables, except where indicated, correspond to the actual number of speakers of the language.

Of the 10 branches of the Tupí family, the Tupí-Guaraní branch is the one mostly studied. Languages of this branch have a higher degree of lexical and morphological similarities to each other when compared to languages of other branches. Internal classification of the Tupí family is currently in the early stages, but what is known about languages outside the Tupí-Guaraní branch allows a few generalizations to be made about Tupian languages as a whole. Larger genetic relations between Proto-Tupí and other families of languages, especially Macro-Jê and Karíb, have been proposed by Greenberg (1987) and Rodrigues (1985b, 1999, 2000) (see **Macro-Jê; Cariban Languages**).

General Properties of Tupian Languages

From the point of view of phonetics and phonology, Tupian languages do not have intricate consonantal and/or vocalic systems. Rodrigues (1999: 112) has reported that consonant systems across the family vary from 10 to 19, and Rodrigues and Dietrich (1997) proposed that Proto-Tupí has a six-vowel system. It is common that languages of various branches have a phonological distinction between short and long vowels (cf., Jurúna and Xipáya, of the Jurúna branch; Mundurukú, of the Mundurukú branch; possibly all languages of the Tuparí branch; all languages of the Mondé branch; and Karitiána, of the Arikém branch). Furthermore, nearly half of the Tupian branches have languages with either a true

tone system (the Mundurukú and Mondé branches and possibly the Tuparí and Jurúna branches) or a pitch-accent system (Arikém and Ramaráma branches). Stress in Tupian languages is predictable, occurring generally in the last syllable of words. Tupian languages also have a syllable structure that typically does not allow consonant codas word-internally, with the exception of the glottal stop and the glottal fricative. Thus, patterns of consonant-vowel-consonant (CVC) and vowel-consonant (VC) occur exclusively word-finally.

From the point of view of morphology, Tupian languages are agglutinative and isolating. Only a few linguistic categories are marked by affixes – for instance, pronominal prefixes, two or three valence-changing prefixes (causative, comitative causative, and detransitivizer or passivizer), modal markers (usually indicative and gerund), and diminutive/augmentative markers. Categories such as number, gender, tense, and aspect are syntactically marked by particles.

Word classes are well established and easily distinguishable from each other on morphological and/or syntactic/semantic bases. Typical word classes are nouns (including pronouns), verbs (transitive, intransitive and, sometimes, uninflected verbs), postpositions, and particles. Adjectives occur in only a few branches (Arikém, Ramaráma, and Mondé). In all other branches, a descriptive verb fulfills the function of ‘attributes’ and ‘properties.’ Core cases, with the possible exception of Tupí-Guaranian languages, are not morphologically marked. Oblique case marking is conveyed by postpositions, in postpositional phrases. Usually, four or five cases are marked (ablative, allative, dative, instrumental, locative), although languages such as Karo have a larger system; Karo has 12 different postpositions that are used to mark the ablative, abessive, adessive, allative, comitative, dative, dispersive, inessive, instrumental, locative, similitive, and circumjunctive cases.

Nouns, with the exception of those for elements of nature, are categorized as either alienable or inalienable. Alienable nouns generally designate manufactured items, kinship terms, animals, and plants, and occur freely in noun phrases. Inalienable nouns include mostly body parts (and, in some languages, kinship terms), and must occur preceded either by a free noun or a personal prefix (or, in some languages, such as Karo, a personal clitic). The occurrence of positional demonstratives, which mark the lying, standing, sitting, and hanging position of the head noun, is common in Tupian languages. Positional demonstratives are found in Mekéns, Karitiána, Mawé, and Mundurukú.

There is a remarkable class of words called ‘ideophones’ in many Tupian languages. Although

Table 2 Classification of the Tupí-Guaraní branch^a

| Subgroup | Language and clusters ^b | Country | Population |
|----------|------------------------------------|---|--------------------|
| I | Ancient Guaraní | Brazil | Extinct |
| | Chiriguano (Avá) | Argentina/Bolivia/Paraguay | 15 000/35 000/2000 |
| | Izoceño | Bolivia | 15 000 |
| | Guayakí | Paraguay | 850 |
| | Kaiwá | Argentina/Brazil/Paraguay | 500/9000/10 000 |
| | Mbyá | Argentina/Brazil | 1000/2000 |
| | Nhandéva | Brazil | 4900 |
| II | Paraguayan Guaraní | Paraguay and border areas of Argentina and Brazil | 4 000 000 |
| | Xetá | Brazil | 3 |
| | Guarayu | Bolivia | 5000 |
| | Sirionó | Bolivia | 650 |
| | Jorá | Bolivia | 5–10 |
| | Lingua Geral Paulista | Brazil | Extinct |
| | Nheengatu | Brazil | 3000 |
| III | Tupí | Brazil | Extinct |
| | Tupinambá | Brazil | Extinct |
| | Avá (Canoeiro) | Brazil | 100 |
| | Asuriní of Tocantins | Brazil | 200 |
| | Guajajára | Brazil | 10 000 |
| | Parakanã | Brazil | 350 |
| | Suruí of Tocantins | Brazil | 150 |
| IV | Tapirapé | Brazil | 200–350 |
| | Tembé | Brazil | 100–200 |
| | Turiwára | Brazil | Extinct |
| | Anambé of Cairarí | Brazil | 20 |
| | Araweté | Brazil | 80 |
| | Ararandewára-Amanajé | Brazil | 200 (extinct) |
| | Asuriní of Xingú | Brazil | 65 |
| V | | Brazil | 70 |
| | | Brazil | ? |
| | Apiaká | Brazil | ? |
| | Kawahib cluster | | |
| | <i>Amondawa</i> | Brazil | 50 |
| | <i>Karipuna</i> | Brazil | 12–15 |
| | <i>Juma</i> | Brazil | 9 |
| VI | <i>Tenharim</i> | Brazil | 260 |
| | <i>Uru-eu-wau-wau</i> | Brazil | 100 |
| | Kayabi | Brazil | 800 |
| | Parintintin | Brazil | 130 |
| | Kamayurá | Brazil | 270 |
| | North of the Amazon | | |
| | <i>Emerillon</i> | French Guiana | 200 |
| VII | <i>Wayampí (Oyampí)</i> | Brazil/French Guiana | 500/650 |
| | <i>Zo'é</i> | Brazil | 140 |
| | South of the Amazon | | |
| | <i>Guajá</i> | Brazil | 350 |
| | <i>Aurê and Aurá</i> | Brazil | 2 |
| | <i>Urubú-Kaapor</i> | Brazil | 500 |
| | | | |

^aData from Rodrigues and Cabral (2002).^bNames in italics indicate language clusters.

their properties are not yet totally understood and/or described, roughly, ideophones have similarities to intransitive verbs, but their phonological, morphological, syntactic, and discourse behaviors are rather different. Ideophones are found in Karo (Ramaráma), Karitiána (Arikém), Mundurukú (Mundurukú), Xipáya (Jurúna), and Kamayurá (Tupí-Guaraní). Languages of the Awetí, Mawé, Mondé, and Tuparí

branches do not have ideophones, but rather have a class of uninflected verbs.

Syntactic characteristics of Tupian languages include a basic subject-object-verb (SOV) order of clause constituents, with fronting of S or O being used as a syntactic device for emphasis or contrast. The occurrence of clause-chaining constructions, whereby a clause is structured of one main verb in

the finite form plus one or more chained verbs in non-finite or unmarked form, is common and is sometimes erroneously interpreted as serial verb constructions (Jensen, 1990). Typically, coreferential intransitive subjects receive special markings in chained clauses (although this does not characterize a switch-reference system), and transitive subjects are absent (zero-anaphora).

Evidentiality is also a widespread phenomenon in all branches of the Tupí family. Unfortunately, this is not yet fully understood and described, with the exception of Karo (Gabas, 1999) and Kamayurá (Seki, 2000). In Karo, the 11 evidentials are grouped in two categories. One grouping refers to the attitude of the speaker toward the proposition conveyed, and the other refers to the source of information. For Kamayurá, Seki (2000: 104) has described the existence of 'interjective particles' that are used to report to the attitude of the speaker toward the information conveyed. Although Seki does not explicitly analyze these particles as being evidentials, they can easily be interpreted as such.

Tupian languages also have systems of noun classification. In two branches, Mundurukú and Karo, a robust classifier system occurs. In Mundurukú, approximately 50 classifiers occur associated with the preceding noun according to their shape. Classifiers in Mundurukú also occur in concordance with other elements in the noun phrase. In Karo, a set of 11 classifiers occurs, relating to the shape (7), arrangement (2), and gender (1) of the preceding noun (the meaning of the 11th classifier remains unknown). Classifiers in Karo also occur, obligatorily, after an adjective, in concordance. Although languages of other branches do not have classifier systems per se, cognates of classifiers in Karo and Mundurukú occur lexicalized in many words throughout the family, usually the classifier for round objects, *raʔ*; the classifier for concave/convex objects, *ka* or *kap*; and the classifier for flat objects, *peʔ*. This suggests that a system of noun classification already existed in the protolanguage, Proto-Tupí.

Bibliography

- Cabral A S A C (1996). 'Algumas evidências lingüísticas de parentesco genético do Jo'é com as línguas Tupí-Guaraní.' *Moara* 4, 47–76.
- Dooley R A (ed.) (1984). *Estudos sobre línguas Tupí do Brasil. Série lingüística*, 11. Brasília: Summer Institute of Linguistics.
- Drude S (2002). 'Fala masculina e feminina em Awetí.' In Rodrigues A D & Cabral A S A C (eds.) *Línguas indígenas Brasileiras: fonologia, gramática e história (atas do i encontro Internacional do Grupo de Trabalho sobre línguas Indígenas da ANPOLL)*, tomo I. Belém: Editora Universitária UFPA. 177–190.
- Gabas N Jr (1988). *Estudo fonológico da língua Karo. Studies in Native American linguistics* 31. Munich: Lincom.
- Gabas N Jr (1999). A grammar of Karo (Tupi, Brazil). Ph.D. diss., University of California, Santa Barbara.
- Gabas N Jr (2000). 'Genetic relationship among the Ramaráma family of the Tupi stock (Brazil).' In van der Voort H & van de Kerke S (eds.) *Indigenous languages of lowland South America*. Leiden: Research School of Asian, African, and Amerindian Studies (CNWS). 71–82.
- Galucio A V M (2001). The morphosyntax of Mekens (Tupi). Ph.D. diss., University of Chicago.
- Galucio A V M (2002). 'Word order and constituent structure in Mekens.' *Revista da ABRALIN (Associação Brasileira de Lingüística)* 1(2), 51–74.
- Greenberg J H (1987). *Language in the Americas*. Stanford: Stanford University Press.
- Melo A A S (1996). 'Genetic affiliation of the language of the Indians Aurê and Aurá.' *Opción – Revista de Ciencias Humanas y Sociales* 19, 67–81.
- Moore D (1994). 'A few aspects of comparative Tupi syntax.' *Revista Latinoamericana de Estudios Etnolingüísticos* 8, 151–162.
- Moore D (1997). 'Estrutura de cláusulas em gavião de Rondônia.' *Boletim da Associação Brasileira de Lingüística (ABRALIN)* 20, 91–105.
- Moore D (1999). 'Tonal system of the Gavião language of Rondônia, Brazil, in Tupian perspective.' In *Proceedings of the Symposium Cross-Linguistics Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics*. Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA). 297–310.
- Moore D (2002). 'Verbos sem Flexão.' In Rodrigues A D & Cabral A S A C (eds.) *Línguas Indígenas Brasileiras: fonologia, gramática e história (atas do i encontro Internacional do Grupo de Trabalho sobre línguas Indígenas da ANPOLL)*, tomo I. Belém: Editora Universitária UFPA. 139–150.
- Moore D & Galucio A V M (1993). 'Reconstruction of proto-Tupari consonants and vowels. Survey of California and other Indian languages.' *Berkeley* 8, 119–137.
- Rodrigues A D (1964). 'Classificação do tronco lingüístico Tupi.' *Revista de Antropologia* 12, 99–104.
- Rodrigues A D (1966). 'Classificação da língua dos Cinta-Larga.' *Revista de Antropologia* 14, 27–30.
- Rodrigues A D (1980). 'Tupinambá e Mundurukú: evidências fonológicas e lexicais de parentesco genético.' *Estudos Lingüísticos* 3, 194–209.
- Rodrigues A D (1985a). 'Relações internas na família lingüística Tupi-Guarani.' *Revista de Antropologia* 27, 33–53.
- Rodrigues A D (1985b). 'Evidence For Tupi-Carib relationships.' In Stark L R (ed.) *South American Indian languages: retrospect and prospect*. Austin: University of Texas Press. 371–404.
- Rodrigues A D (1999a). 'Macro-Jê.' In Dixon R M W & Aikhenvald A (eds.) *The Amazonian languages*. Cambridge: Cambridge University Press. 164–206.
- Rodrigues A D (1999b). 'Tupí.' In Dixon R M W & Aikhenvald A (eds.) *The Amazonian languages*. Cambridge University Press. 107–124.

- Rodrigues A D (2000). 'Gê-Pano-Carib × Jê-Tupí-Karib: sobre relaciones lingüísticas prehistóricas en Sudamérica.' In *Actas del I Congreso de lenguas Indígenas de Sudamérica*. Lima: Universidad Ricardo Palma 1. 95–104.
- Rodrigues A D & Cabral A S A C (2002). 'Reverendo a classificação interna da família Tupi-Guaraní.' In *Atas do I Encontro Internacional do Grupo de Trabalho de Línguas Indígenas da Associação*

Nacional de Pós-Graduação em Letras e Linguística (ANPOLL), tomo I. Belém: Editora Universitária UFPA. 327–337.

- Rodrigues A D & Dietrich W (1997). 'On the linguistic relationship between Mawé and Tupi-Guarani.' *Diachronica* 14, 265–304.
- Storto L R (1999). Aspects of a Karitiana grammar. Ph.D. diss., MIT.

Turkic Languages

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Development and Classification

The Turkic language family was first attested in 8th century inscriptions. Turkic-speaking groups first appeared in the Inner Eurasian steppes, from where they moved to Central Asia, Eastern Europe, the Middle East, Siberia, etc. Because of their high mobility, Turkic expanded over a huge area.

The Proto-Turkic network of varieties was dissolved by an early split of Oghur or Bulgar Turkic. Its modern representative, Chuvash, a descendant of Volga Bulgar, differs from Common Turkic by specific phonetic representations, e.g., *r* and *l* instead of *z* and *š* in words such as *šér* 'hundred' and *śul* 'year' (Turkish *yüz* 'hundred,' *yaş* 'age'). A second split is represented by Khalaj, which retains a reflex of Proto-Turkic **p*- as *h*-, e.g., *hadaq* 'foot.' Dialect splitting has led to further differentiation of Common Turkic. There is no mutual intelligibility throughout the family today. The following division combines the current areal distribution with genealogical and typological features.

1. The Southwestern or Oghuz branch contains a western subgroup comprising Turkish, Gagauz, and Azerbaijanian (Azerbaijani, Northern and Azerbaijanian, Southern), a southern subgroup comprising dialects of southern Iran and Afghanistan, and an eastern subgroup comprising Turkmen and Khorasan Turkic.
2. The Northwestern or Kipchak branch has a western subgroup comprising Kumyk, Karachay-Balkar, Crimean Tatar, and Karaim, a northern subgroup comprising Tatar and Bashkir, and a southern subgroup comprising Kazakh, Karakalpak, Kipchak Uzbek, Nogai, and Kirghiz (of different origin, but strongly influenced by Kazakh).

3. The Southeastern or Uyghur-Karluk branch has a western Uzbek subgroup and an eastern Uyghur subgroup.
4. The Northeastern or Siberian branch has a southern heterogeneous subgroup comprising Sayan Turkic (Tuvan, Tofan), Abakan (Yenisei) Turkic (Khakas, Shor), Chulym Turkic, Altai Turkic (Altai, Northern and Southern), and a northern subgroup comprising Yakut (Sakha) and Dolgan.
5. Chuvash is geographically situated in the north-western area (Volga region).
6. Khalaj is geographically situated in the southwestern area (central Iran).

Deviant languages in China are Salar, of Oghuz origin, Yellow Uyghur (Yugur, West) and Fu-yü (Manchuria), both of south Siberian origin.

One traditional classificatory criterion is the final consonant of the word for 'nine.' Its representation as *r* in Chuvash *täxxär* separates Oghur from Common Turkic (Turkish *dokuz*). The intervocalic consonant in the word for 'foot' divides most Northeastern languages, Chuvash, Khalaj, etc. from the rest, which exhibits -*y*- (Turkish *ayak*), e.g., Tuvan *adaq*, Khakas *azax*, Chuvash *ura*. Oghuz Turkic differs from the rest by loss of suffix-initial velars, e.g., *qal-an* [remain-PART] instead of *qal-yan* [remain-PART] 'remaining.' Final -*G* is devoiced in the Southeast (Uyghur *tay-liq* [mountain-DER] 'mountainous'), preserved in southern Siberia (Tuvan *day-lïy* [mountain-DER]), and lost elsewhere (Turkish *dağ-lı* [mountain-DER]).

Most older linguistic stages are insufficiently known. Written sources, where available, provide no direct information on spoken varieties. Early Oghuz and Bulgar (East Europe, 6th–7th centuries) are unknown. There are no texts in the language of the Khazars (7th–10th centuries). Pecheneg and Kuman, predecessors of West Kipchak, are only known from loanwords, titles, and names.

Written Varieties

Turkic literary varieties have emerged in various cultural centers. Many older Turkic empires, however, used foreign languages for administration (Sogdian, Persian). Muslim Turks often used Persian for poetry, and Arabic for religious and scientific writing. Russian has played an important role for many groups. The following main stages of written Turkic may be distinguished.

1. An older pre-Islamic East Old Turkic period (8th century–), is represented in inscriptions, manuscripts, and block prints. East Old Turkic proper is documented in stone inscriptions (Orkhon Valley), which celebrate the rulers of the Second Eastern Türk Empire, in other inscriptions found in Mongolia and the Yenisei and Talas valleys, and also in a few manuscripts. The Old Kirghiz inscriptions are of this type. Old Uyghur is first recorded in the period of Uyghur rule over the Eastern Empire. Early Old Uyghur is attested in runiform inscriptions and manuscripts. From the 10th century on, Old Uyghur became the medium of a flourishing literary culture in the Tianshan-Tarim area, attested in texts of Buddhist, Manichaean, and Nestorian content.
2. A middle Turkic period comprises various early Islamic varieties.

The first East Turkic written language, Karakhanid (11th century–), developed in Kashgar, is close to Old Uyghur but lexically influenced by Arabic and Persian. Maḥmūd of Kashgar provides information (1073) on Karakhanid and other contemporary Turkic varieties.

Khorezmian Turkic, used in the 13th–14th centuries in the Golden Horde and Mamluk Egypt, is based on the older languages but contains Oghuz and Kipchak elements.

This tradition is continued in Chaghatay (15th century–). Early Chaghatay contains regional elements of the Timurid area. Later, Chaghatay became the dominant written language of Central Asia, eventually conquering an immense area of validity and developing regional varieties.

The first West Turkic written language is Volga Bulgar, insufficiently known from epitaphs of the 13th and 14th centuries. Information on early Kipchak Turkic is given in the *Codex Cumanicus*, compiled by Christians, and in dictionaries and grammars written in Mamluk Egypt and Syria.

Oghuz Turkic is first represented by Old Anatolian Turkish (13th century–), which was a subordinate written medium until the end of Seljuk rule. Old Ottoman is the initial stage of Ottoman, which

begins with the foundation of the Ottoman Empire in 1307. In Azerbaijan a literary language developed from the 15th century on.

3. A premodern period (16th century–) begins with the development of regionally influenced written languages. Middle and Late Ottoman became the leading written language with an abundantly rich literature. Chaghatay continued to play a major role and remained the literary language of all non-Oghuz Muslim Turks until a century ago.
4. A modern period begins in the second half of the 19th century with the formation of regional written languages. The political division of the Turkic-speaking world in the 20th century and the language policies pursued in the Soviet Union, Turkey, China, and Iran had dramatic effects that increasingly obstructed transregional linguistic contacts. A dozen ‘national’ languages with a narrow radius of validity emerged. In Turkey, Ottoman was replaced by modern Turkish. The social importance of many Turkic languages was very limited. After the recent political developments, their significance is rapidly increasing, but the varieties spoken in Iran, Afghanistan, Iraq, etc., still have poor possibilities to develop.

Various scripts and script systems have been applied to Turkic. A specific runiform script was created for Old East Turkic. Most Old Uyghur texts are written in Uyghur script, originating in the Near East and later taken over by Mongols and Manchus. It is similar to the Sogdian script, which is also used in Buddhist texts. A few Buddhist manuscripts are written in Brahmi script, Manichaean texts in Manichaean script, and Nestorian texts in Syriac script. Arabic script was used for the languages of the Islamic era (still used in China for Uyghur and Kazakh). A unified Roman-based script was introduced for several languages in the early Soviet period, but later replaced by different Cyrillic-based scripts. A Roman-based alphabet was introduced in Turkey in 1923. Most of the newly established Turkic republics have introduced or are introducing Roman-based scripts.

Contacts

The massive displacements of Turkic-speaking groups throughout their history have led to various phenomena induced by contacts with Iranian, Slavic, Mongolic, Uralic, etc. Speakers of Turkic have copied lexical, phonetic, morphological, and syntactic elements, whereas non-Turkic (e.g., Iranian, Greek, Finno-Ugric, Samoyedic, Yeniseian, Tungusic) groups shifting to Turkic have exerted substrate influence by

copying native elements into their new varieties. Languages such as Chuvash, Yakut, Salar, Yellow Uyghur, Khalaj, Karaim, and Fu-yü have long developed in isolation from their relatives, preserving old features and acquiring new ones in their environments. Long and intense interaction with Iranian in Central Asia, Iran, Afghanistan, etc., has led to profound convergence phenomena. Massive foreign influence has sometimes caused considerable typological deviations, e.g., drastic structural changes in Karaim and Gagauz under Slavic impact.

Most written languages have been strongly influenced by Persian and Arabic. In Chaghatay (Chagatai) and Ottoman, lexical borrowing contributed to a remarkable richness of the vocabularies, whereas grammar was much less affected. The overload of Persian and Arabic in Ottoman led to strong puristic efforts in the 20th century to create a so-called Pure Turkish.

Internal convergence processes have resulted in leveling of languages of the central area. Several Turkic koinés have been used as transregional codes for trade and intergroup communication, e.g., Azerbaijani in Iran and the Caucasus region.

Linguistic Features

Despite their huge area of distribution, Turkic languages share essential phonological, morphological, and syntactic features.

They have a synthetic word structure with numerous highly applicable derivational and grammatical suffixes, and a juxtaposing technique with clear-cut morpheme boundaries and predictable allomorphs. These agglutinative principles yield considerable morphological regularity and transparency. Exceptions include traces of vowel gradation in the pronominal declination, e.g., Turkish *ben* 'I,' *ban-a* [I-DAT] 'to me.' The agglutinative structure is partly deranged in languages of the northeast and southeast. Some languages, e.g., Uzbek, even display borrowed prefixes.

The syllable contains minimally a vowel with maximally one preceding and one subsequent consonant. Vowel hiatus and consonant clusters are avoided.

Most languages exhibit eight short vowel phonemes, *a, i, o, u, e, ɛ, ö, ü*, classified according to the features front vs. back, unrounded vs. rounded, and high vs. low. Proto-Turkic long vowel phonemes are preserved in Turkmen, Yakut, and Khalaj. Iranian and Slavic phonetic influence has sometimes affected the front vs. back distinctions. Tatar, Bashkir, Chuvash, and Uyghur exhibit systematic vowel shifts. Chuvash, Gagauz, Karaim, etc., have developed palatalized

consonants, e.g., Karaim *méní* 'I'. Tuvan and Tofan exhibit a glottal element signaling strong obstruents, e.g., *aʔt* 'horse' vs. *at* 'name.'

The most general sound harmony phenomenon is an intrasyllabic front vs. back assimilation. An inter-syllabic front vs. back harmony causes neutralization of the front vs. back distinction under the influence of the preceding syllable. If applied consistently, it excludes back and front syllables in a word, e.g., Turkish *ev-ler-im-e* [house-PL-POSS.1.SG-DAT] 'to my houses,' *at-lar-ım-a* [horse-PL-POSS.1.SG-DAT] 'to my horses.' Some languages only display this kind of harmony, whereas others also apply a rounded vs. unrounded harmony, neutralization of the distinction rounded vs. unrounded in high suffix vowels, e.g., Turkish *el-im* [hand-POSS.1.SG] 'my hand,' *gül-üm* [rose-POSS.1.SG] 'my rose.' Languages such as Yakut and Kirghiz apply this harmony to low-vowel suffixes as well, e.g., *börö-lör* [wolf-PL] 'wolves.' There are numerous exceptions to harmony rules in loanwords. Further allomorphs are created by various consonant assimilations.

The rules of word accent vary. A high pitch accent, interacting with a dynamic stress accent, mostly falls on the last accentable syllable of native words.

The morphological structure has remained relatively stable through the centuries. The main word classes are nominals (nouns, adjectives, pronouns, numerals) and verbals. The primary stems can be used as free forms, e.g., *at* 'horse,' *at!* 'throw!'. From verbal and nominal stems, which are sharply distinguished, expanded stems are formed. Nominals take plural, possessive, case, and specific derivational suffixes. Grammatical gender is not marked. The verbal morphology comprises markers of actionality, voice, possibility, negation, aspect, mood, evidentiality, tense, person, interrogation, etc. Voice is expressed by passive, reflexive-middle, causative, and cooperative-reciprocal suffixes. The order and combinability of suffixes is basically common to all Turkic languages.

Constructions with postposed auxiliary verbs (post-verbs) express actional modifications. A few constructions have developed into aspect-tense categories, e.g., Turkish *gel-iyor* [come-PRES] 'comes' < **gel-e yorî-r* [come-CONV run-AOR] ('runs coming'). Possibility markers are formed with auxiliary verbs such as *bil-* 'to know' and *al-* 'to take,' e.g., Kirghiz *ber-e al-* [give-CONVAUX.POTEN] 'to be able to give.'

Turkic languages share many syntactic characteristics. With respect to relational typology, they adhere to the nominative-accusative pattern. They have a head-final constituent order, with dependents preceding their heads. The unmarked order of clause constituents is subject + object + predicate (SOV).

Adjectival, genitival, and participial attributes precede the head of the nominal phrase. Postpositions are used instead of prepositions. There is no agreement in number or case between dependents and heads. The focus position is in front of the predicate core. The unmarked constituent order is often deviated from for discourse-pragmatic reasons. Contact-induced word order changes are common, e.g., in Gagauz, which has become an SVO language.

Preposed subordinate clauses are based on verbal nouns, participles, and converbs. The use of postposed subordinative patterns with conjunctions are typical effects of Iranian and Slavic influence. Most languages possess conjunctions, even coordinative ones meaning 'and,' 'or,' and 'but' of Persian, Arabic, or Russian origin.

Turkic lacks definite articles. The indefinite article is formally identical with the numeral 'one' Genitival attributes, expressing a possessor, stand in the genitive, whereas their head, indicating a possessed entity, carries a possessive suffix, e.g., Turkish *at-ın baş-ı* [horse-GEN head-POSS.3.SG] 'the head of the horse.' The dominant type of nominal compounds follows the

pattern noun + noun + possessive suffix, e.g., Turkish *el çanta-sı* [hand bag-POSS.3.SG] 'handbag.'

All Turkic varieties exhibit numerous loanwords. Arabic and Persian loans are frequent in all Islamic-Turkic languages. The Iranian influence is strong in Uyghur, Uzbek, and varieties of Iran and Afghanistan. Many languages have been subject to considerable Mongolic and Slavic influence. Loans and calques from European languages have become increasingly important. The Turkic languages spoken in China exhibit old and recent loans from Chinese.

Bibliography

- Deny J *et al.* (eds.) (1959). *Philologiae turcicae fundamenta*, 1. Aquis Mattiacis: Steiner.
- Johanson L (2002). *Structural factors in Turkic language contacts*. London: Curzon.
- Johanson L & Csátó É Á (eds.) (1998). *The Turkic languages*. London: Routledge.
- Menges K H (1995). *The Turkic languages and peoples. An introduction to Turkic studies*. Wiesbaden: Harrassowitz.
- Róna-Tas A (1991). *An introduction to Turkology*. Szeged: University of Szeged.

Turkish

R Underhill, San Diego State University, San Diego, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Turkish (natively *Türkçe*), the official language of the Republic of Turkey, is spoken by a large proportion of the Turkish population. There are also Turkish speakers in the Balkans, particularly in Greece, Bulgaria, and the former Yugoslavia, although there has been extensive population inflow from those countries into Turkey, and there is a substantial minority of Turkish speakers in Cyprus. There are Turkish-influenced Turkic dialects in Iraq in the region of Kirkuk, where the speakers are called Turkmen or Turkomans. The *Ethnologue* entry for Turkish gives a population of roughly 46 million speakers in Turkey, and 61 million in all countries.

Turkish belongs to the southwestern, or Oghuz (*Oğuz*), group of Turkic languages. This group also includes Azerbaijani, spoken in Azerbaijan and in adjacent areas of Iran; Qashqay and related dialects, spoken in the Zagros mountain area of Iran; Türkmen, spoken in Turkmenistan; and Gagauz,

spoken in Bulgaria, in Romania, and principally in Moldova, although there has been substantial migration from Moldova to Turkey. Central Asian Turkic languages include the national languages of Kazakhstan, Uzbekistan, and Kyrgyzstan, and a number of others. Turkic, in turn, belongs to the Altaic family of languages, which also includes the Mongol and Manchu-Tunguz language families. Though this relationship has recently been called into question, it was proved convincingly by Poppe more than a generation ago (Poppe, 1960). Wider affinities of the Altaic family have been suggested for Korean, and even for Japanese.

Turkish scholars divide the history of the Turkish language into three periods: (1) Old Anatolian Turkish (*Eski Anadolu Türkçesi*), comprising texts dating from the earliest arrival of Turkic speakers in Anatolia, through the Seljuk period to the formation of the Ottoman Empire; (2) Ottoman (*Osmanlıca*), the language of the Ottoman Empire, heavily influenced by Arabic and Persian; and (3) Modern Turkish (*Yeni Türkçe*), dating from the overthrow of the Ottoman Empire and from the Turkish language reform movement of the 1920s and 1930s. The Turkish language reform movement was

launched by Atatürk as part of his overall plan to distance Turkey from Middle Eastern, specifically Arabic and Persian, influences, in favor of European influence. This movement in the language area included most noticeably the replacement of the Arabic writing system with a Latin alphabet in 1928, and a drive to replace Arabic and Persian vocabulary, once pervasive in Ottoman texts, with vocabulary drawn or constructed from Turkish sources, or Turkish-looking inventions. The drive to cleanse the lexicon has waxed and waned over the interim and has acquired political correlates: writers on the left tend to use neologisms; those on the right use a more traditional vocabulary. There has been no corresponding attempt to rid the lexicon of European or English terminology (for more on the language reform movement, see Lewis (1999)). In 1997, a committee of the American Association of Teachers of Turkic Languages attempted to create a standardized English terminology for Turkish, which is used here.

Phonology

Phonemes

Consonants The International Phonetic Association (IPA) representations of the Turkish consonant system are shown in Table 1. Turkish uses 21 letters for consonants: b c ç d f g ğ h j k l m n p r s ş t v y z. These represent the expected sounds, except as follows:

| Letter | Sound |
|--------|-------|
| c | [dʒ] |
| ç | [tʃ] |
| j | [ʒ] |
| ş | [ʃ] |

In the following discussions, [tʃ] and [dʒ] will henceforth be written /ç/ and /j/, since they function in all phonological respects as members of the natural class of stops, not as clusters. The letters *k g l* each stand for two sounds: a plain velar or lateral [k g l] and a front velar or palatal [c ʝ ʎ]. In words of Turkish origin, the front velar variant occurs with front vowels and the plain velar occurs with back vowels.

Table 1 International Phonetic Association symbols for Turkish consonants

| Labial | Dental | Palatal | Front velar | Velar | Glottal |
|--------|--------|---------|-------------|-------|---------|
| p | t | tʃ | c | k | |
| b | d | dʒ | ʝ | g | |
| f | s | ʃ | | | h |
| v | z | ʒ | | | |
| m | n | | | | |
| | l | ʎ | | | |
| | r | ʕ | | | |

In words of Arabic origin, however, /c ʝ ʎ/ can occur with back vowels, giving rise to pairs and thus distinctive contrasts, as in *kar* ‘snow’ [kar] and *kâr* ‘profit’ [car].

The letter ğ, or *yumuşak ge* ‘soft g’, has no consonantal sound. It normally represents an historical or underlying /g/ that has been deleted; in some Anatolian dialects, it survives as a voiced fricative [ɣ]. Most commonly, ğ lengthens the preceding vowel in syllable-final (coda) position, and represents nothing between vowels, as in *dağ* ‘mountain’ [da:] and *dağa* ‘mountain (dat)’ [daa].

Vowels Turkish vowels are traditionally represented in a ‘cube’ shape, consisting of all possible values of the features, front/back, high/low, and rounded/unrounded, as in Figure 1. Each vowel can occur long, from the deletion of ğ, and the vowels /e i a u/ can occur long in Arabic loanwords, giving a total of 16 vowel phonemes. The vowel letters are for the most part self-explanatory, except for *ı*, an undotted ‘i,’ which is a high back unrounded vowel, IPA [u]. All Turkish vowels are phonetically lax, except sometimes before *y* or *ğ*, thus *a e i ı o ö u ü* sound like [a ε ɪ u ɔ œ ʊ ʏ]. Because the difference between *ı* and *i* is distinctive, it must be maintained for capitals also, i.e., *I* and *İ*.

Stress Stress in Turkish consists of higher pitch, rather than greater loudness on the accented syllable. Stress is normally on the last syllable of the word; as affixes are added, stress moves rightward:

- (1) *él* ‘hand’
ellér ‘hands’
ellerim ‘my hands’

There are a number of exceptions to final stress. Some words have inherent nonfinal stress, and in these cases stress does not move with the addition of affixes. Inherently stressed words include most loans, which have their own rule for accent; in such cases, the accent may fall on a syllable other

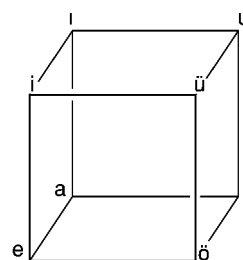


Figure 1 Turkish vowels. Front vowels are represented at the front of the cube, high vowels are at the top, and rounded vowels are to the right. Reproduced from Underhill R (1976) Turkish grammar. Cambridge: MIT Press. With kind permission by MIT Press.

than that which is stressed in the source language, as in *sinéma* ‘cinema’ and *Kenédi* ‘Kennedy’. Some affixes are prestressing; stress then falls on the preceding syllable and remains there as additional affixes are added. The rules for stress and much else in Turkish phonology are extensively worked out in Demircan (2001).

Phonological Rules

Turkish being an agglutinating language, suffixes are added to stems in such a manner that segmentation is relatively easy. However, a number of changes take place in both stems and suffixes when this happens.

Vowel Harmony Vowel harmony involves the two features front/back and rounded/unrounded. It is a syllable-to-syllable process by which each vowel conditions the following vowel, according to the following rules:

1. Any of the vowels can occur in the first syllable of a word.
2. A noninitial vowel assimilates to the previous vowel in frontness.
3. A noninitial high vowel assimilates to the previous vowel in rounding. A noninitial low vowel is unrounded. Thus /o ö/ do not appear in harmonic suffixes.

The process is illustrated in Table 2, which shows how the stem, dative (suffix *-yA*), and objective (suffix *-yI*) case forms of a set of nouns are used (in morphophonemic transcription, the symbol *A* represents the alternation between /a/ and /e/, and the symbol *I* represents the alternation /i ı u ü/). A few native words and very many foreign words are nonharmonic, such as *kardeş* ‘brother’, *otel* ‘hotel’, and *sigorta* ‘insurance’. This has led some scholars to claim that vowel harmony no longer holds for stems (Clements and Sezer, 1982). In the case of a nonharmonic word, suffixes are controlled by the last syllable, as in *asansör* ‘elevator’ (plural *asansörler*) and *kredikart* ‘credit card’ (plural *kredikartlar*).

Table 2 Turkish vowel harmony

| Stem | Gloss | Dative | Objective |
|------------|------------|-------------|-------------|
| <i>bal</i> | ‘honey’ | <i>bala</i> | <i>balı</i> |
| <i>kıl</i> | ‘hair’ | <i>kıla</i> | <i>kılı</i> |
| <i>ok</i> | ‘arrow’ | <i>oka</i> | <i>oku</i> |
| <i>buz</i> | ‘ice’ | <i>buza</i> | <i>buzu</i> |
| <i>ev</i> | ‘house’ | <i>eve</i> | <i>evi</i> |
| <i>il</i> | ‘province’ | <i>ile</i> | <i>ili</i> |
| <i>göl</i> | ‘lake’ | <i>göle</i> | <i>gölü</i> |
| <i>gül</i> | ‘rose’ | <i>güle</i> | <i>gülü</i> |

Other Phonological Rules Beyond vowel harmony, stems and suffixes have a highly changeable nature. Suffix-initial voiced stops devoice after a stem ending in an unvoiced consonant. Many suffixes have different postconsonantal and postvocalic forms. Stems also undergo a number of rules designed to maintain canonical syllable structure, particularly in closed syllables. Among the rules applying to syllables are final devoicing, epenthesis, degemination, and vowel shortening. There are many details concerning these rules, but as an extreme example, the verbal noun suffix best written as *-DIg* has 16 forms:

-dik/dık/duk/dük/tık/tuk/tük/diğ/dığ/duğ/düğ/tiğ/tığ/tüğ/tüg

Morphology

Turkish is an agglutinating language in which suffixes, in some cases a large number of them (the lists of suffixes in the following sections are not exhaustive), are added fairly transparently to stems:

- (2) *ev* ‘house’
evler ‘houses’
evlerim ‘my houses’
evlerimiz ‘our houses’
evlerimizde ‘in our houses’
evlerimizdeki ‘which is in our houses’

The Noun Paradigm

Noun stems may have the following inflectional suffixes, in order:

1. Plural *-lAr* (as in *baba* ‘father’, *babalar* ‘fathers’ and *deve* ‘camel’, *develer* ‘camels’).
2. Possessive (possessed agreement).
3. Case (as in *oda* ‘room’).

- (3) Nominative: *oda*
 Genitive (*-(n)In*): *odanın*
 Dative (*-yA*): *odaya*
 Objective (*-yI*): *odayı*
 Locative (*-DA*): *odada*
 Ablative (*-DAn*): *odadan*
 Instrumental/comitative (*-y-IA*): *odayla*

The Verb Paradigm

Starting with the verb root, a number of derivational suffixes can be added to build up the verb stem. These include reflexive, reciprocal, causative, passive, impossibility, negative, and abilitative forms. At this point, from the verb stem, it is possible to go in a number of directions. For a finite (‘tensed’) verb, the next step is a tense suffix, followed normally by a personal ending:

| | | |
|----------------------|------------|-----------------------|
| (4) General present: | gelirim | 'I come', 'I'll come' |
| Progressive: | geliyorum | 'I am coming' |
| (Definite) past: | geldim | 'I came' |
| Unwitnessed past: | gelmişim | 'I (supposedly) came' |
| Future: | geleceğim | 'I will come' |
| Necessitative: | gelmeliyim | 'I ought to come' |
| Optative: | geleyim | 'let me come' |
| Conditional: | gelsem | 'if I come' |

There is also a wide range of nonfinite suffixes possible at this point for the formation of subordinate clauses. These include verbal nouns or nominalizations, participles, and adverbial clause suffixes (traditional 'converbs').

Auxiliary Suffixes

Finally, there is a group of suffixes that can be categorized under the heading of 'auxiliary'. They can be added both to verbal and nonverbal predicates, hence a separate auxiliary category. They include most prominently the personal endings, but also some morphemes that can be called 'aspects', although they are not all aspects any more than the tenses are all tenses (abbreviations: SG, singular; PROG, progressive):

- (5) Yorgun -du -m.
tired -PAST -1SG
'I was tired'.
- (6) Gel -iyor -du -m.
come -PROG -PAST -1SG
'I was coming'.

The aspects are past *-y-DI*, dubitative *-y-mİş*, and conditional *-y-sA*. Furthermore, there is an adverbial aspect *-y-ken*. These look very similar to some tenses, i.e., definite past *-DI*, unwitnessed past *-mİş*, and conditional *-sA*, but they differ in morphology, meaning, and prosody (all auxiliary suffixes are prestressing).

The inferential/quotative, sometimes called dubitative (DUB), *-y-mİş*, deserves special discussion. This aspect, and to some extent the corresponding tense, *-mİş*, are used when the speaker wishes to be disassociated from the truth of the utterance – for example, when the speaker has information that has only been heard or recently found out (VB, verb):

- (7) Sen tembel -miş -sin.
you lazy -DUB -2SG
'They say you are lazy'.
- (8) Geçen sene hasta -lan-mış-sın.
past year sick -VB-DUB-2SG
'(I heard) you got sick last year'.

The dubitative can also be used for statements for which the speaker does have personal knowledge of

the fact, but is expressing something unexpected or surprising – for example, after trying a food that the speaker had expected to dislike:

- (9) Bu yemek iyi -miş!
this food good -DUB
'This food is good!'

Syntax

Unmarked (normal) word order is subject-object-verb, as shown in the following example (OBJ, objective; DAT, dative):

- (10) Hasan mektub -u
Hasan letter -OBJ
Ayşe-ye gönder -di.
Ayşe-DAT send -PAST
'Hasan sent the letter to Ayşe'.

However, this is complicated by the fact that Turkish has pragmatically conditioned word order, by which the information status of noun phrases, rather than their grammatical function, determines their placement in the sentence. Many of the basic principles were worked out by Erguvanlı (1984). The topic is sentence initial; thus, any of the terms of Example (10) could be initial, depending on whether Hasan, the letter, or Ayşe is the topic. New information comes in the preverbal position, thus any of the terms of Example (10), if indefinite, would move preverbally:

- (11) Mektub-u Ayşe-ye bir
letter-OBJ Ayşe-DAT a
arkadaş gönder-di.
friend send-PAST
'A friend sent the letter to Ayşe'.

In fact, preverbal position is focus position; thus, *wh*-words are found here, as well as words questioned contrastively, the focused words in the answers to *wh*-questions, or any focused argument. Though the canonical sentence pattern for English might be written as subject-verb-object-X, where X is everything else, the pattern for Turkish would be topic-X-focus verb, and is thus determined by pragmatic rather than by grammatical conditions. Furthermore, sentences are not necessarily verb final. Backgrounded or unstressed information can move to the right of the verb, producing what is traditionally called a *devrik cümle* (*tümce*), or 'inverted sentence' (NEG, negative; PL, plural):

- (12) Ver-me çocuğ-a kibrit-ler-i.
give-NEG child-DAT match-PL-OBJ
'Don't give the child the matches'.

The focus in Example (12) is 'don't give,' and the child and the matches will have been previously

mentioned or are clear in the context, i.e., are ‘given’ in the sense of functional syntax.

Turkish is a left-branching and head-final language in which nouns follow adjectives (Example (13)), possessives (Example (14)), and relative clauses (Example (15)); postpositions follow noun phrases (Example (16)), and verbs follow direct objects, even subordinate clauses (Example (17)) (GEN, genitive; POSS, possessive; LOC, locative; PART, participle; ABL, ablative; VN, verbal noun; FUT, future):

- (13) çok küçük bir çocuk.
very small a child
‘A very small child’.

- (14) Enver-in şapka-sı.
Enver-GEN hat-POSS
‘Enver’s hat’.

- (15) Köşe-de otur-an kız.
corner-LOC sit-PART girl
‘The girl who is sitting in the corner’.

- (16) Bu haber-den dolayı.
this news-ABL because
‘Because of this news’.

- (17) Hasan-in yarın
Hasan-GEN tomorrow
gel-eceğ-in-i duy-du-m.
come-VN.FUT-3SG-OBJ hear-PAST-1SG
‘I heard that Hasan will come tomorrow’.

Notice from Example (17) that Turkish is a pro-drop language (‘pronoun dropping’; i.e., subject pronouns

normally are not used, as in Latin or Spanish). Overt pronouns appear in cases of focus or contrast, including topic change. Because relative clauses precede head nouns, and direct objects (including noun complement clauses) precede the main verb, Turkish sentences sometimes give the impression of having the reverse word order from English. English speakers reading Turkish sometimes find it easier to start at the end of a sentence and read toward the front, and Turkish speakers report that they do the same in reading English.

Bibliography

- Clements G N & Sezer E (1982). ‘Vowel and consonant disharmony in Turkish.’ In van der Hulst H & Smith N (eds.) *The structure of phonological representations, part II*. Dordrecht: Foris. 213–255.
- Demircan Ö (2001). *Türkçenin ses dizimi*. İstanbul: Der Yayınları.
- Erguvanlı (Taylan) E (1984). *The function of word order in Turkish grammar*. Berkeley: University of California Press.
- Kornfilt J (1997). *Turkish grammar*. London: Routledge.
- Lewis G (1999). *The Turkish language reform: a catastrophic success*. Oxford: Oxford University Press.
- Poppe N (1960). *Vergleichende grammatik der Altaischen sprachen*. Wiesbaden: Harrassowitz.
- Taylan E E (2001). *The verb in Turkish*. Amsterdam: John Benjamins.
- Underhill R (1976). *Turkish grammar*. Cambridge: MIT Press.

Turkmen

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Turkmen (*türkmen dili*, *türkmençe*) belongs to the southwestern or Oghuz branch of the Turkic language family, which also includes Turkish. It is mainly spoken in Turkmenistan (*Türkmenistan döwleti*), which is located in the Transcaspien region and whose capital is Ashgabat. Turkmenistan borders on Iran and Afghanistan in the south, Uzbekistan in the east, and Kazakhstan in the north. The area of distribution of Turkmen extends from the southeastern shore of the Caspian Sea to the Kazakh-speaking area in the north, the Karakalpak-speaking area in the north-east, the Uzbek-speaking area in the east, beyond

the Amudarya River, and the Persian (Farsi, Western) and Khorasan Oghuz (Khorasani Turkish) areas in the south, beyond the borders to Afghanistan and Iran. Though Turkmen make up 85% of the 4.8 million inhabitants, only 72% speak Turkmen. The other main languages are Russian (12%) and Uzbek (9%). Turkmen-speaking groups also live in the Russian Federation, Kazakhstan, Tajikistan, China, etc. The total number of speakers amounts to nearly 5 million.

The designation ‘Turkmen’ is not unequivocal. Older Oghuz varieties spoken in Khorezm, Khorasan, Azerbaijan, Anatolia, and other regions in the Near East were referred to as ‘Turkmen.’ Several nomadic groups in Anatolia, Iraq, etc. are still called ‘Turkmen’ without being Turkmen in a linguistic sense.

Since the mid-1990s, language policy aims at consolidating Turkmen as the state language and to remove the Russian dominance. Turkmen is gaining

more social functions. The 1992 constitution defines it as the “official language of inter-ethnic communication.” Geographic names and administrative terms have been changed from Russian to Turkmen. In practice, however, Russian has maintained its importance in most spheres of public communication.

Origin and History

The Turkmen go back to the Turkic-speaking Oghuz confederation of tribes, whose Inner Asian steppe empire collapsed in 744. Certain Oghuz groups migrated into the region between the Syrdarya and Ural rivers. By the late 10th century, the Seljuk dynasty was founded, and an autonomous state was established on the lower Syrdarya. The Seljuks left this region in the middle of the 11th century and migrated westwards. Their modern descendants are the Turks of Khorasan, Azerbaijan, and Turkey. The speakers of Turkmen are mainly descendants of non-Seljuk groups that did not take part in these migrations.

During the Mongol conquests in the 13th century, the remaining Oghuz tribes were pushed into the Karakum desert and the region east of the Caspian Sea. From the 16th century on, Turkmen groups migrated to Khorezm, to the southern part of today's Turkmenistan, and to Khorasan, absorbing local Turkic and Iranian elements. The major migrations of the Salır, Ersari, Sarıq and Teke tribes took place in the 17th century. In the 18th century, the Turkmen conquered the whole core area that they inhabit today.

Most tribes were subsequently divided and controlled by the Uzbek khanates of Khiwa and Bukhara, while the Persian shahs tried to subdue the southern tribes. The dependence of Khiwa and Persia came to an end after the mid-19th century. Some decades later, Russia annexed the Turkmen territory, which caused many Turkmen groups to emigrate to Afghanistan and Iran. The Turkmen area was first administered as the Trans-Caspian district in the Governorate of Turkistan. In 1924, Turkmenistan was proclaimed a Socialist Soviet Republic.

In connection with the dissolution of the Soviet Union, Turkmenistan declared its sovereignty in 1990, achieved its independence in 1991 (after a popular referendum), and adopted its new constitution in 1992.

Related Languages and Language Contacts

The closest relative of Turkmen is Khorasan Turkic (Khorasani), spoken in northeastern Iran and Khorezm, a distinct language with which it constitutes the eastern subbranch of Oghuz. Azerbaijanian

(Azerbaijani) and Turkish represent the western subbranch. The specific features of Turkmen are partly archaic and partly innovative, due to language contact. Within the Turkic family, Khorasan Turkic, Uzbek, and Karakalpak are the most important contact languages. Turkmen has had intensive contacts with Persian and, during the last century, with Russian.

The Written Language

Old Turkmen is not clearly documented in written sources. The oldest records of ‘Turkmen’ relate to Oghuz varieties in general. Oghuz texts of the following centuries do not exhibit any specific Turkmen features. A written Turkmen literature began in the 18th century, but the language used is a variety of the classical Chaghatay (Chagatai) language. A Turkmen standard language was created in the Soviet era and formed mainly from 1928 on. It was based on the Teke dialect as spoken in the Ashgabat region.

Arabic script was used in the first period. Two script reforms, in 1922 and 1925, aimed at reflecting spoken features more adequately. A Roman-based alphabet that reflected most of these features rather accurately was in use from 1928 to 1940. A variant of the Cyrillic alphabet was adopted in 1939–1940. Since the early 1990s, there has been a transition to a Roman-based script again. In 1993, the final version of a Roman-based alphabet was adopted to replace the Cyrillic one. It has several unique letters that distinguish it from Turkey's alphabet and the newly adopted alphabets of other Turkic republics.

Distinctive Features

Turkmen exhibits most linguistic features typical of the Turkic family (*see Turkic Languages*). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation, e.g., *A = a/e*, *I = i/i*. Segments in round brackets only occur after consonant final stems. Hyphens are used here to indicate morpheme boundaries.

Phonology

Turkmen has, like Yakut and Khalaj, preserved Proto-Turkic long vowels in a consistent way, e.g., *a:t* ‘name’ < *a:t* (but at ‘horse’ < *at*), *dö:rt* ‘four’ vs. Turkish *dört* < *tö:rt*. The orthography does not normally mark vowel length, but *ü:* in words of Turkic origin is expressed by *üy*, e.g., *süyt* for [θü:t] ‘milk’.

Proto-Turkic *e*: is mostly represented by Turkmen *i*, which mostly corresponds to Azerbaijani *e*, e.g., *gi:č* 'late' (Azerbaijani *geç*, Turkish *geç*).

A striking feature of Turkmen pronunciation is the presence of the interdental fricatives *θ* and *ð*, which correspond to *s* and *z* in other Turkic languages, e.g., *θið* 'you' (Turkish *siz*).

As in Azerbaijani, the word-initial back velar *g*-corresponds to *q*- in other Turkic languages, e.g., *g̈i:ð* 'girl' (Azerbaijani *giz*, Turkish *kız*). Initial *b*- is preserved in *ber*- 'to give', *bar*- 'existing', *bar*- 'to go' and *bol*- 'to become' (Turkish *ver*-, *var*, *var*-, *ol*-). The bilabial fricative *β* is used instead of labiodental *v*, e.g., *a:β* 'hunt' (Turkish *av*). It appears as the glide *w* between two vowels or between a liquid and a vowel. The bilabial fricative *f* is frequently replaced by the stop *p* in loans, e.g., *pikir* 'thought' (Turkish *fikir*).

Suffix vowels mostly assimilate to the quality of the preceding vowel. Turkmen displays both front vs. back harmony and rounded vs. unrounded harmony. The latter also includes suffixes with low vowels, e.g., *toy-do* [feast-LOC] 'at the feast' vs. *öy-dö* [house-LOC] 'in the house'. Long *a*: and *e*: are not rounded; there are also other exceptions. Though the orthography represents the vowels of rather closely, rounding harmony is not consistently represented. Rounding is only expressed in high vowels and not beyond the second syllable. The tendency towards rounded low suffix vowels is also observed in languages such as Kirghiz, Altay Turkic (Altai), and Yakut.

Numerous consonant assimilations are observed, e.g., *men-ne* [I-LOC] 'in me', *g̈ið-dan* [girl-ABL] 'from the girl', *yol-loş* [way-DER] 'comrade' (Turkish *ben-de* [I-LOC], *kız-dan* [girl-ABL], *yol-daş* [way-DER]). They are mostly not reflected in the orthography.

In copies of loanwords, nonpermissible consonant clusters are dissolved by means of prothetic or epenthetic vowels, e.g., *uθθul* 'chair' < Russian *stul*, *pikir* 'thought' < Arabic *fikr*. In recent loanwords from Russian, these vowels are not reflected orthographically.

Grammar

The comparative degree of adjectives is formed with *-rA:K*, e.g., *kičire:k* [small-COMP] 'smaller, rather small' (*kiči* 'small'). The demonstrative pronouns *bu*-, *šu*-, *ol* and *şo[l]* form a fourfold deictic system, expressing various degrees of distance (Turkish *bu*, *o* and *şu*).

The old present tense, mostly called 'indefinite future,' is formed with *-Ar*, e.g., *bil-er* [know-AOR] 'will know' (Turkish *bil-ir* [know-AOR]), *oqa:r*

[read-AOR] 'will read' (Turkish *oku-r* [read-AOR]). The negative marker is *-mAð* in the third person, and *-mAr* in the other persons, e.g., *gel-mer-in* [come-NEG.AOR-1.SG] 'I will not come' (Turkish *gel-me-m* [come-NEG.AOR-1.SG], Azerbaijani *gel-mer-em* [come-NEG.AOR-1.SG]). A more focused present tense is formed with *-yA:r*, often contracted to *-yA*, e.g., *bil-ye:r* [know-PRES.3.SG] 'knows', *oqa-ya:r* [read-PRES.3.SG] 'reads, is reading'. A few verbs exhibit contracted forms without this marker: *du:r* [stand-PRES.3.SG] 'is standing', *oẗi:r* [sit-PRES.3.SG] 'is sitting', *yaẗi:r* [lie-PRES.3.SG] 'is lying'. These forms can be used with a converb marker to express a continuous present, e.g., *oqa:-p oẗi:r* [read-CONV AUX-PRES.3.SG] 'is reading'.

The second-person imperatives include an unmarked singular, e.g., *gel* [come.IMP.2.SG] 'come!', a form expressing insistence, e.g., *gel-gin* [come-IMP.2.SG], a plural form, e.g., *gel-iñ* [come.IMP.2.PL], and intensifying forms, e.g., *gel-θen-e* [come-IMP.2.SG] (singular) and *gel-θe-ñið-lä:ñ* [come-IMP.2.PL] (plural). The first-person plural has a special form that only refers to the speaker and the addressee, e.g., *gel-eli-ñ* [come-IMP.1.PL] 'let us come', *gel-eli* [come-IMP.1.INCL] 'let us come (you and me)'.

The future marker *-jAK* and the intentional marker *-mAK-čI* lack personal markers, e.g., *men gel-jek* [I come-FUT] 'I will come' (Turkish *gel-eceğ-im* [come-FUT.1.SG]), *men yað-maq-č̈i* [I write-INTENT] 'I intend to write'.

Turkmen has a postterminal ('past') participle marker *-An* and an intraterminal ('present') participle marker *-yA:n*, e.g., *bil-en* [know-POSTTERMINAL.PART] 'having known', *bil-ye:n* [know-INTRATERMINAL.PART] 'knowing'. A categorical negation is formed with the participle in *-An* + possessive suffix + *yo:q* 'non-existing', e.g., *al-am-o:q* (<*al-an-im yo:q*) [take-POSTTERMINAL.PART-POSS.1.SG not-existing] 'I did/do not take at all'. There is a postterminal converb marker *-(I)p*, e.g., *oyno-p* [play-POSTTERMINAL.CONV] 'having played'. The corresponding marker of Turkish and Azerbaijani displays the uncontracted form *-(y)Ip/-(y)Ib*, e.g., Turkish *oyna-yip* [play-POSTTERMINAL.CONV].

Among the evidential markers, the inflectional suffix *-(I)p-dlr*, negated *-mAn-dlr*, forms an evidential (indirective) past, e.g., *gel-ip-dir* [come-POSTTERMINAL.CONV-EV.3.SG] 'has evidently come'. The copula particle *eken* combines with various participles, e.g., *gel-en eken* [come-POSTTERMINAL.PART EV.PARTICLE] 'has obviously arrived'. The copula particle *-mIş* suggests second-hand information, e.g., *gel-ip-miş-in* [come-POSTTERMINAL.

CONV-EV.3SG] 'has reportedly come'. A presumptive intraterminal (present, imperfect) is formed with *-yA:n-dlr*, a presumptive postterminal (perfect) with *-A:n-dlr*, e.g., *bar-ya:n-nür* [go-INTRATERMINAL.PART-PRESUMP.3.SG] 'is probably going', *bar-an-nür* [go-POSTTERMINAL.PART-PRESUMP.3.SG] 'has probably gone'.

A number of postverb constructions with converbs plus auxiliary verbs, *goy-* 'to put', *git-* 'to go away', *çiq-* 'to go out', *dur-* 'stand', *otur-* 'sit', *yör-* 'move', etc., express modifications of the manner in which the action denoted by the lexical verb is carried out.

Lexicon

The Turkmen vocabulary is basically of southwestern Turkic origin, though it also contains words typical of the Northwestern and Southeastern branches of Turkic. There are synonyms representing Oghuz and non-Oghuz types, e.g., *gapî* and *işik* 'door', *dodaq* and *erîn* 'lip'. The vocabulary contains numerous words of Arabic and Persian origin, borrowed from Persian and representing the traditional sphere of Islamic civilization, e.g., *xat* 'letter', *înθa:n* 'human being', *ša:t* 'glad', *gül* 'flower', *ireňk* 'color'. The Turkmen conjunctions are mainly of Arabo-Persian origin, e.g., *we* 'and', *emma:* 'but'. Words of Russian origin, borrowed from the 19th century on, represent phenomena of modern life, e.g., *poθyolok* 'settlement', *ğaḏyet* 'newspaper', *fe:rma* 'farm'. The vocabulary contains many recent internationalisms borrowed via Russian.

Dialects

Turkmen dialects and subdialects are referred to by the names of tribes and clans. One main dialect group comprises the Teke, Yomud, Sariq, Salir, Gökleng and Ersari dialects, which are rather close to Standard Turkmen. The Teke dialect, occupying the central

area, has two subdialects, Mari and Akhal, the latter spoken in the Ashgabat region. The Yomud dialect is spoken on the southeast shore of the Caspian Sea and in the northern part of Turkmenistan. Ersari dialects are spoken in the eastern part of the country. The second main dialect group is found in the regions on and beyond the borders to Iran and Uzbekistan. These dialects are more distant from Standard Turkmen, lacking, for example, the interdental pronunciation of the sibilants *s* and *z*.

An isolated variety of Turkmen is Türkpen (Russian Trukhmen), spoken by small groups (ca. 12 000) on the lower Kuma River in the Stavropol region of Northern Caucasus. Türkpen is strongly influenced by Noghay (Nogai). Its speakers are descended from Turkmen tribes that migrated here in the 18th century from the Mangyshlak region east of the Caspian Sea. Salar, spoken in western China, seems to go back to an early Turkmen variety.

Bibliography

- Baskakov N A *et al.* (eds.) (1970). *Grammatika turkmenskogo jazyka*. Aşxabad: Akademija Nauk Turkmenskoy SSR.
- Bazin L (1959). 'Le turkmène.' In Deny J *et al.* (eds.) *Philologiae turcicae fundamenta 1*. Aquis Mattiacis: Steiner. 308–317.
- Clark L (1998). *Turkmen grammar*. *Turcologica* 34. Wiesbaden: Harrassowitz.
- Dulling G K (1960). *An introduction to the Turkmen language*. Oxford: Oxford University Press.
- Hanser O (1977). *Turkmen manual. Descriptive grammar of contemporary literary Turkmen. Texts. Glossary*. Wien: Verlag des Verbandes der wissenschaftlichen Gesellschaften Österreichs.
- Johanson L (2001). 'Turkmen.' In Garry J & Rubino C (eds.) *Facts about the world's major languages: An encyclopedia of the world's major languages, past and present*. New York, Dublin: The H. W. Wilson Company, New England Publishing Associates. 766–769.

This page intentionally left blank

U

Ugaritic

J A Hackett, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

The Ugaritic language was rediscovered after a 3000-year gap, when, in spring of 1928, a farmer discovered a tomb at Minet el-Beida, on the Mediterranean, in what is now Syria, about 12 km from modern Latakia and a few hundred yards away from a large tell called Ras Shamra, 'Cape Fennel.' In 1929, the French began excavating what turned out to be a large necropolis. They soon moved on to the nearby tell of Ras Shamra, and in May of that year the first Ugaritic cuneiform clay tablets were found. After the first tablets were published in 1930, it was clear that the repertoire of signs was small (only 30), and so the writing system was assumed to be alphabetic and without vowels. Within months, the language was essentially deciphered. The identification of the site with ancient Ugarit was confirmed in the early 1930s with the discovery of a tablet that mentioned Niqmaddu, king of Ugarit.

Ugaritic is one branch of the Northwest Semitic languages, along with the Canaanite languages, the several forms of Aramaic, and other less well-documented languages. It is written in a cuneiform alphabet on clay tablets. Because the acrophonic linear alphabet predates the Ugaritic alphabet by several centuries, Ugaritic cuneiform was probably devised to adapt the idea of the alphabet to the medium of clay and stylus. Our earliest abecedaries, which are texts that list the letters of an alphabet written in a standard order, come from Ugarit, and they exhibit both the usual West Semitic order and, strikingly, the South Semitic order in a very few texts. Why abecedaries in this South Semitic order were present at Ugarit is so far unknown.

Ugaritic exhibits individual signs for 27 consonants of the West Semitic languages, plus three extra signs. There are two extra 'aleph signs, plus one for a sibilant that is used for loanwords. The three 'aleph signs are transcribed 'a, 'i, and 'u: 'a is used when an 'aleph in a word is followed by the vowel /a/, 'i is used when 'aleph is followed by /i/ or /e/ (<*ay), and

'u is used when 'aleph is followed by /u/ or /o/ (<*aw). A syllable-closing 'aleph is marked by 'i. These three signs have been very helpful in determining the vocalization of Ugaritic words, as have syllabaries that include Ugaritic words spelled out in Akkadian cuneiform, which is syllabic and so includes vowels. The Ugaritic consonants, given in the indigenous alphabet order, are, ' b, g, ḫ, d, h, w, z, ḥ, t, y, k, š, l, m, ḏ, n, z, s, ʾ, p, š, q, r, ṭ, g, ṭ (plus, as was noted above, two extra ' signs and a sibilant sign used for loanwords). The vowels reconstructed for Ugaritic are a, i, u, ā, ī, ū, o (<*aw), and e (<*ay). This cuneiform alphabet also exists in a shorter form of 22 signs, indicating that where this shorter alphabet is used, several mergers of consonants have taken place. A few tablets written in this shorter alphabet come from the site of Ugarit, but many were found farther south, at Sarepta and Kamid el-Loz in Lebanon, and at Taanach, Mt. Tabor, and Beth Shemesh in Israel.

The city-state of Ugarit was an important port, with its position on the Mediterranean and its proximity to Cyprus on the west, and its access to inland routes to the north and east. There are writings found at Ugarit in several different languages: besides Ugaritic, there are texts in Akkadian (the lingua franca of the time), Sumerian, Hittite (both syllabic cuneiform and hieroglyphic), Egyptian, Hurrian, and Cypri-Minoan. Texts found at the sites of Mari, Alalakh, and Amarna, among others, mention the city-state. The Ugaritic texts cover a short period of time, probably late 14th to early 12th century B.C. Excavation continues, but as of this writing, approximately 50 poetic texts and 1500 prose texts have been found at Ras Shamra and at neighboring Ras Ibn Hani. The poetic texts are mythological; the prose texts are ritual and other cultic texts, administrative documents, letters, omens, medical texts, and school exercises. The poetic mythological texts are characterized by parallelism, as in these couplets from the Baal myth:

Sea sends messengers/Judge River, a delegation;
Message of Sea, your master/your lord, Judge River.

Like other West Semitic languages, Ugaritic has prefix- and suffix-conjugation verbs, *yaqtulu/qatala*,

but there are in addition two more prefix-conjugations: *yaqtul* and *yaqtula*. The prefix-conjugation *yaqtul* serves as both a jussive, or indirect imperative, and as a preterit. The prefix-conjugation *yaqtula* is less well understood, but appears to serve as a volitive form; it also, however, seems to occur in subordinate (especially purpose) clauses. The verb stems that are extant in Ugaritic are G, Gt, D, tD, N, Š (a causative), Št (reflexive of the causative).

Nominals in Ugaritic have masculine and feminine gender and singular, dual, and plural number. There are three cases in the singular – nominative, genitive, and accusative; the plural is diptotic – nominative and oblique. Nouns occur in both absolute (unbound) and bound states. The bound state is used for initial members of genitive chains called construct chains (see **Semitic Languages**) and for nouns before pronominal possessive suffixes. There is no marked definite article in Ugaritic. There is evidence for -a-insertion in the plurals of nouns of the shape C₁vC₂C₃–: the (nominative) plural is C₁vC₂aC₃ūma. For example, ‘king’ (nominative) is *malku*, and ‘kings’ is *malakūma* (we can compare Biblical Hebrew *mélek*, plural *mēlākîm*).

Bibliography

Daniels P T & Bright W (eds.) (1996). *The world's writing systems*. New York: Oxford University Press.

- Day P L (2002). ‘Ugaritic.’ In Kaltner J & McKenzie S L (eds.) *Beyond Babel: a handbook for biblical Hebrew and related languages*. Atlanta: Society of Biblical Literature Press. 223–241.
- Ginsberg H L (1970). ‘The Northwest Semitic languages.’ In Mazar B (ed.) *The world history of the Jewish people*, vol. 2. Givatayim: Jewish History Publications. 102–124.
- Goetze A (1941). ‘Is Ugaritic a Canaanite dialect?’ *Language* 17, 127–138.
- Gordon C (1965). *Ugaritic textbook*. Rome: Pontifical Biblical Institute.
- Pardee D (2004). ‘Ugaritic.’ In Woodard R (ed.) *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press. 288–318.
- Pitard W (1999). ‘The alphabetic Ugaritic tablets.’ In Watson & Wyatt (eds.). 46–57.
- Rainey A F (1996). ‘Who is a Canaanite? A review of the textual evidence.’ *Bulletin of the American Schools of Oriental Research* 304, 1–15.
- Smith M S (2001). *Untold stories: the Bible and Ugaritic studies in the twentieth century*. Peabody, MA: Hendrickson.
- Tropper J (1994). ‘Is Ugaritic a Canaanite Language?’ In Brooke G J (ed.) *Ugarit and the Bible*. Münster: Ugarit-Verlag. 343–353.
- Watson W G E & Wyatt N (eds.) (1999). *Handbook of Ugaritic studies. Handbuch der Orientalistik* 39. Leiden: Brill.

Ukrainian

S Young, University of Maryland Baltimore County, Baltimore, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Ukrainian, with some 36 million speakers in the Ukrainian Republic, forms with Russian and Belorussian the East Slavic branch of the Slavic language family. The standard language, which is written in the Cyrillic alphabet, has its roots in the 19th century – no enduring literary tradition had been able to form before this time – and is based on the relatively recent and uniform southeastern dialect. Since the late 19th century, the West Ukrainian (Galician) speech of L'viv has also played a role in the formation of the national standard.

Among the vocalic features that distinguish Ukrainian from the rest of East Slavic are the preservation of *o* in unstressed syllables: *vodá* /vodá/ ‘water’ (Rus., BR /vadá/), and the merger of East Slavic (ESL.) *i* with *y*

to give a central-front mid vowel (represented in transliteration by *y*): *synij* ‘blue,’ like *syn* ‘son’ (Rus. *sinij* : *syn*). A new *i* developed in turn from ESL. **ē*: *lis* /līs/ ‘woods’ (Rus., BR /līs/) and from *e* and *o* in a secondarily closed syllable: *šist* ‘six’ (gen. *šestý*), *nis* ‘nose’ (gen. *nósa*). *e* > *o* after hushers and *j*: *čotýry* ‘four,’ *johó* ‘his’ (Rus. *četýre*, *jegó*), but *těplyj* ‘warm’ (Rus. *těplyj* /t'ò-).

In contrast to Russian and Belorussian, Ukrainian consonants are not palatalized before *e* or *y* (the merger of ESL. *i* and *y*): *nestý* ‘to carry’ (Rus. *nestí* [-tj]), but there is palatalization before the new *i* representing ESL. **ē*, *e*, *o*: *dity* [dī-] ‘children’ (Rus. *děti*), *nis* -nī- ‘nose’ (Rus. *nos*). Stem-final *c* is typically palatalized: *kinec* [-tsj] ‘end,’ gen. *kincjá* (Rus. *konéc*, *koncá*); final labials lose palatalization: *hólub* ‘dove’ (Rus. *gólub*). Common Slavic /g/ has become /h/. Like Belorussian, Ukrainian has *w* (written *v*) corresponding to Russian *v* in a closed syllable: *právda* [právda] ‘truth,’ and in some cases (including

the masculine past tense marker) to *l*: *vovk* [vowk] 'wolf' (Rus. *volk*), *buw* [buw] 'was, masc.' (fem. *bulá*). Unlike other East Slavic languages, there is no regressive devoicing of voiced consonants: *kázka* 'tale' (with *z* preserved), or final devoicing: *did* 'grandfather' (with final *d*).

In addition to the six nominal case forms of Russian and Belorussian, Ukrainian has a regular vocative (*sýnu* 'son!', nom. *syn*). As in Belorussian, there is an alternation of velar and dental stems in certain case forms: nom. *rik* 'year,' loc. *róci*; nom. *rih* 'corner,' loc. *rózi*. The verb has two regular conjugation patterns, illustrated by *nestý* 'to carry' (I) and *xodýty* 'to walk, go' (II): 1SG *nesú*, *xodžú*, 2SG *neséš*, *xódyš*, 3SG *nesé*, *xódyt*, 1PL *nesemó*, *xódym*, 2PL *neseté*, *xódyte*, 3PL *nesút*, *xódjat* (like Belorussian, but unlike Russian, the 3rd person ending is palatalized). Unlike Russian or Belorussian, there is no alternation of velar and palatal stems in 1st conjugation verbs, the palatal stem having been generalized: *mohtý* 'to be able': *móžu*, *móžeš* (BR *mahú*, *móžaš*).

Lexically, Ukrainian lacks the Church Slavisms characteristic of Russian (Ukr. *skoročú* 'shorten.1SG

PF,' with ESL. *s-*, *oro*, *č*; cf. Rus. *sokraščú*, with ChSl. *so-*, *ra*, and *šč*), but shows a large number of borrowings from Polish: *cikávyy* 'interesting' (Pol. *ciekawyy*, but Rus. *interesnyj*), *raxúnok* 'bill, account' (Pol. *rachunek*, but Rus. *sčët*), *otrymáty* 'to receive' (Pol. *otrzymać*, but Rus. *polučít*).

Bibliography

- Pugh S M & Press I (1999). *Ukrainian: a comprehensive grammar*. New York and London: Routledge.
- Shevelov G (1966). *Die ukrainische Schriftsprache 1798–1965*. Wiesbaden: O. Harrassowitz.
- Shevelov G (1979). *A historical phonology of the Ukrainian language*. Heidelberg: C. Winter.
- Shevelov G (1980). 'Ukrainian.' In Schenker A & Stankiewicz E (eds.) *The Slavic literary languages: formation and development*. New Haven: Yale Concilium on International and Area Studies. 143–160.
- Shevelov G (1993). 'Ukrainian.' In Comrie B & Corbett G (eds.) *The Slavonic languages*. London and New York: Routledge. 947–998.

United States of America: Language Situation

D Sharma, King's College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

The linguistic landscape of the United States, though dominated by English, encompasses an unusual diversity of indigenous and immigrant languages. No federal law currently grants English the status of official language, but it is used for virtually all official and institutional functions. Americans tend to be relatively monolingual in English (82% in 2000, **Figure 1**), and Spanish (11%, **Figure 1**) and other languages (7%, **Figure 2**) have a minority status in terms of size of speech community and institutional support (**Figure 2**).

American English

Regional and Social Varieties

English was first established in America by permanent settlers in Jamestown, Virginia, in 1607. By 1780, the number of people of European and African origin had increased to 2.8 million but more than 20% of European Americans were still from non-English-speaking communities, predominantly German, Dutch,

Swedish, Irish, and French. This heterogeneity influenced the lexical stock of American English (e.g., *bayou*, *caribou*, *prairie* (French); *cookie*, *waffle* (Dutch); *noodle*, *snorkel* (German); *corral*, *ranch* (Spanish)) as well as its regional dialect features; Minnesota English, for instance, bears traces of Swedish phonology and syntax. German (German, Standard) once had a substantial presence, but native use is now primarily limited to the dialect of German known as Pennsylvania Dutch.

Standard American English is distinctive in its phonology (rhoticity, except in parts of the South and the Northeast; greater use of /æ/, e.g., *fast*, *can't*; intervocalic flapping of /t/, e.g., *butter*, *writer*; widespread leveling of the vowel distinction in *caught* and *cot*, except in the Northeast), syntax (simple past in perfect contexts, e.g., *Did you see that film yet?*; use of *gotten*), lexicon (*sidewalk*, *carpark*, *elevator*, *schmuck*), and spelling (*center*, *neighbor*, *analyze*; Noah Webster's *American Dictionary of the English Language* [1828] introduced many revisions). Early linguistic atlases (Kurath, 1949) used isoglosses of lexical variants such as *pail/bucket* to identify three primary English dialect divisions in the United States – South, North and, to a lesser extent, Midland – within which further minor dialect divisions occur. More

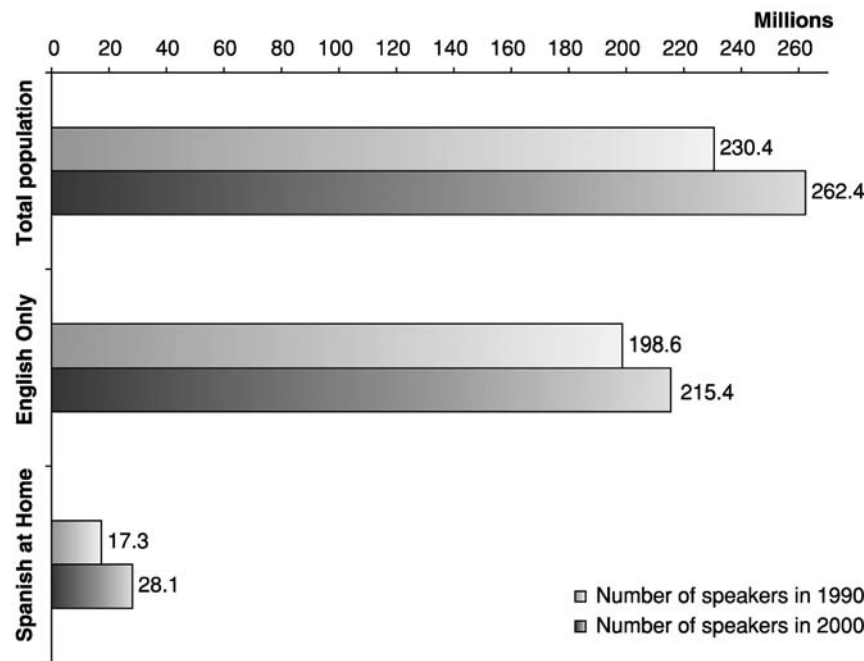


Figure 1 Use of English and Spanish relative to total population in 1999 and 2000 (population 5 years and over). Source: Data from U.S. Bureau of the Census (2003).

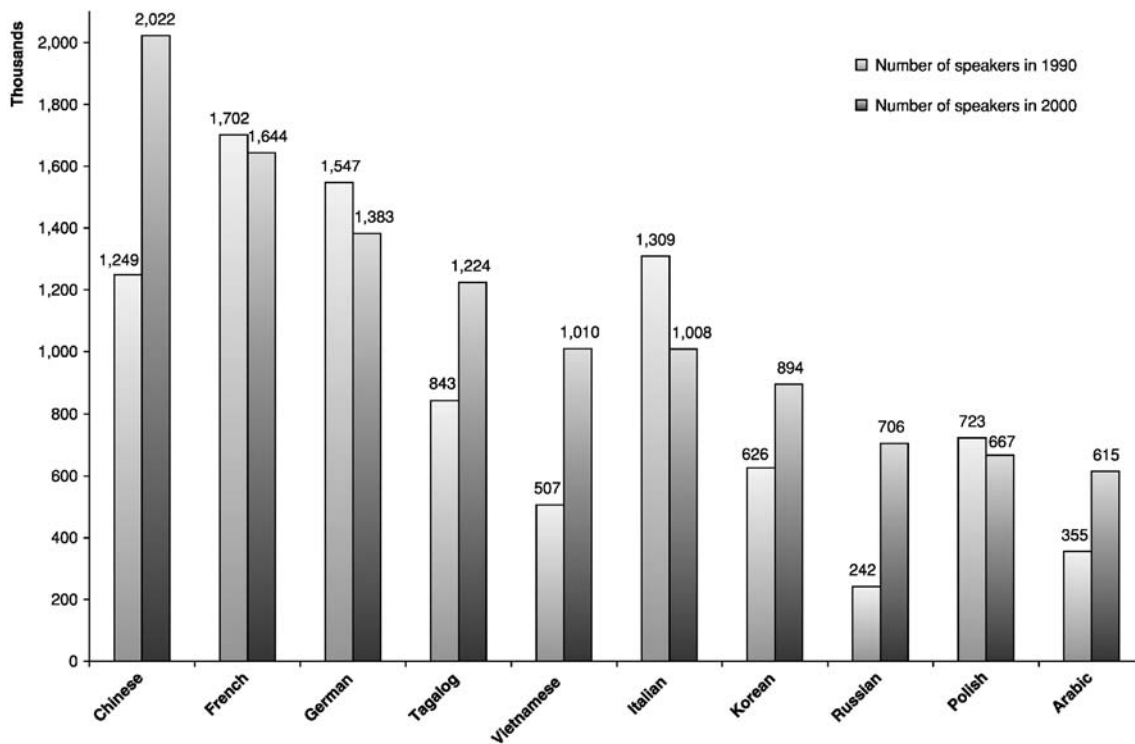


Figure 2 Ten Languages most frequently spoken at home other than English and Spanish in 1999 and 2000 (population 5 years and over). Source: Data from U.S. Bureau of the Census (2003).

recent studies of contemporary dialectal phonological systems continue to reflect these divisions; new dialects are now also beginning to coalesce in more recently settled parts of the West.

Early English-speaking settlers arrived from distinct dialect regions of England and as the frontier later shifted westward, their distinct speech patterns spread along conduits of travel (see **Figure 3**). While

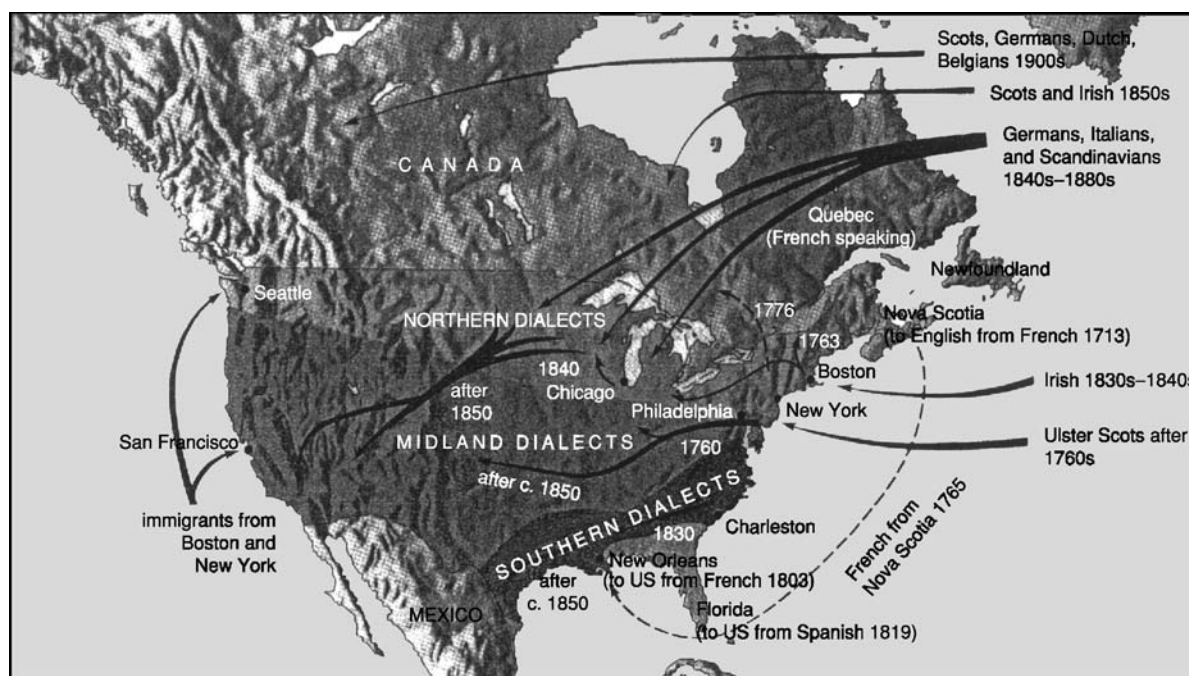


Figure 3 European settlement of North America since the mid-eighteenth century. Source: Reproduced from Graddol, Leith, and Swann (1996: 199).

certain features of American English have been argued to originate in the dialects of Early Modern English that first came to America (e.g., rhoticity; use of /æ/; *gotten*; *mad* 'angry'; *fall* 'autumn'), most dialect distinctions were rapidly leveled through early admixture in settlements. The distinctive features of present-day American English dialects therefore tend to derive more from ongoing language change than from early British English.

Pioneering work by William Labov and other sociolinguists, beginning in the 1960s, has demonstrated that social groupings are also a key factor in American dialects. For instance, the Northern Cities Vowel shift – a series of shifts in pronunciation in the area encompassing Detroit, Chicago, Buffalo, and Cleveland – results in certain linguistic features that function as social markers of class, ethnicity, age, and gender, and the associated prestige or stigma of such markers effects dialect change. This research has also indicated that despite the influence of media certain dialect boundaries, e.g., the North–South division, are strengthening in some respects.

African-American English

The variety spoken by many African Americans bears several defining linguistic features in its phonology (word-final consonant cluster simplification, e.g., *told*, *best*; use of /t, d, f, v/ for /θ, ð/, e.g., in *these*, *with*, *thumb*, *bath*), syntax (invariant *be* for habitual

meaning; nonstandard auxiliary use of *been* and *done*; null copula, e.g., *He workin'*; negative inversion and multiple negation, e.g., *Ain't nobody told me nothing.*), lexicon, and styles of discourse (e.g., toasting, signifying, playing the dozens). Research has shown these features to be systematic and rule-governed, as in all dialects. British English and Creoles have both been proposed as possible origins.

In 1996, the linguistic status of African-American English came under public scrutiny as the Oakland School Board in California passed a resolution declaring a social and educational need to recognize that what they termed Ebonics was the primary language of many students in the county. Although the Linguistic Society of America passed a resolution affirming the importance of recognizing African-American Vernacular English as a systematic dialect, the intensity of the public debate surrounding the school board's resolution led to its ultimate dissolution. The controversy unmasked deeply opposed popular views on the cultural status of vernacular dialects.

The Debate over Bilingualism

Early supporters of installing English as the official language of the United States included Benjamin Franklin and Noah Webster, and the English-Only movement continues this effort. As of 2004, 23 states have adopted Official English laws. However, many

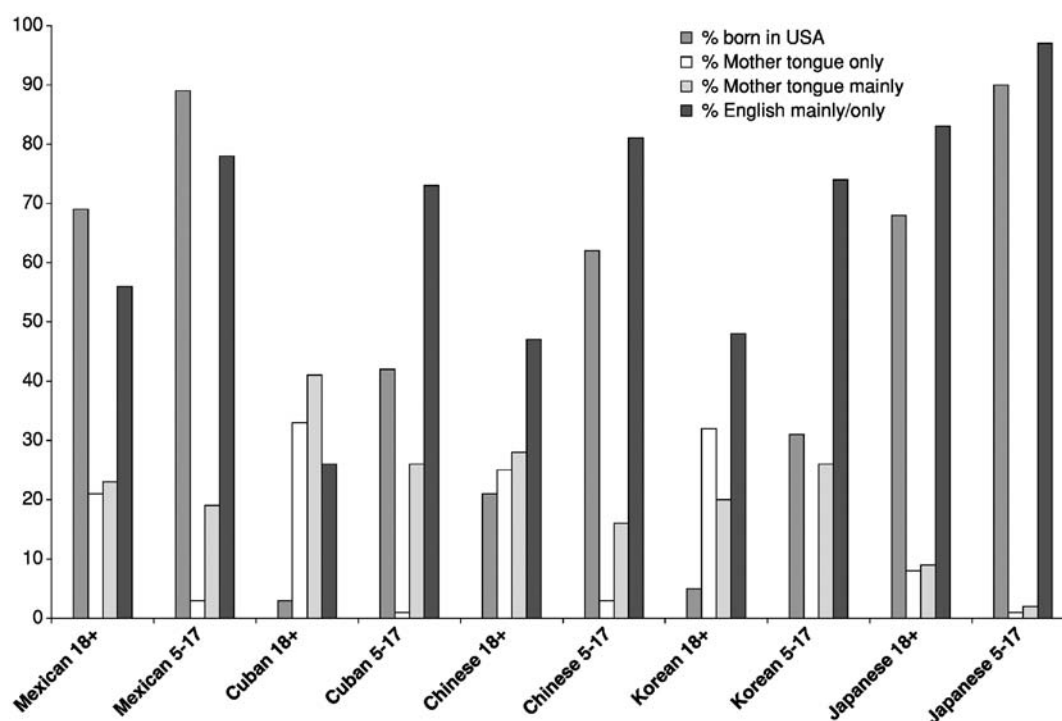


Figure 4 Language use and nativeness across generations among selected immigrant groups. Source: Data from López (1982) as discussed by R. Bayley in Finegan and Rickford (2004: 274).

English-Only claims, e.g., immigrant resistance to learning English and detrimental effects of bilingualism, have been discredited: research finds consistently high rates of language shift to English among immigrants (see Figure 4), and the popular belief during the first half of the 20th century that bilingualism was detrimental to intellectual development has received no empirical support. In 1968, the Title VII Bilingual Education Act allocated federal funds to children with special linguistic needs. The Official English movement resists measures of this sort, while the English Plus movement, advocating a more bilingual model for the United States, supports them.

Spanish in the United States

The arrival of Spanish in the United States predates that of English, and its development in the Southwest and the Northeast has followed distinct historical and demographic patterns.

Spanish colonization began in Florida with Juan Ponce de León's visit in 1513, and spread soon after to Louisiana and the Southwest, where it was administered by the Spanish Viceroyalty, with colonial Spanish remaining the local prestige language for almost two centuries. After the Mexican-American war, almost half of Mexico was ceded to the United States in 1848, including all of present-day California, Nevada, and Utah and parts of Texas, New Mexico,

Colorado, Arizona, and Wyoming. Sustained Mexican migration has continually reinforced the Spanish-speaking population of many of these states. The Southwestern states are now home to just under half of the Spanish-speaking population in the United States. Sometimes termed Chicano Spanish, the Southwestern variety bears characteristics of Mexican Spanish and American English. English influence can be seen in lexical innovations (e.g., *libreria* (not *biblioteca*) 'library'; *parientes* (not *padres*) 'parents'; *puchar* 'to push'; *fensa* 'fence'; *cama king* 'king-size bed'); English-based phonology (e.g., *moven* for *mueven*, 'they move') and syntax (phrasal constructions in place of complex morphology) are also common.

Spanish in the Northeast primarily originates from Puerto Rico, the Dominican Republic, Cuba, and Colombia. In 1898, after the Spanish-American war, Puerto Rico became a territory of the United States and was the first major source of Spanish-speaking immigration to the East Coast. The majority of other immigrants arrived later; Cuban refugee migration, for instance, rose dramatically after the 1959 coup. While some phonological traits of these varieties are shared, such as deletion or aspiration of syllable-final /s/, other regional distinctions may persist: e.g., dropping of syllable-final // and /r/ (Cuban) and raspy velar /r/ (Puerto Rican). As colonial Spanish developed first in the Caribbean, the Northeastern United States varieties have brought many Native

American, African, and Creole loans into American English, e.g., *canoe* (Native American), *banana* (African), *bodega* (Caribbean Spanish).

Due to extensive language shift to English, a continuum of societal bilingualism has emerged in Hispanic communities, ranging from fluency in Spanish to symbolic use of Spanish by English-dominant bilinguals. Alongside the influence of English on Spanish structure, this bilingualism has given rise, particularly among English-dominant bilinguals in the younger generation, to ‘Spanglish,’ a hybrid style consisting of proficient and sustained code-switching between Spanish and English. Chicano English, by contrast, is a variety of English with Spanish influence.

Indigenous Languages

Native American Languages

The languages indigenous to America have undergone extensive decimation through contact with sociopolitically empowered colonial languages. Legislation punishing instruction or use of native languages and mandating English as the exclusive language of instruction was enforced in Indian reservations from the 19th century. Estimates place the number of native languages at the time of European contact at 300–600; the current figure stands at approximately 175, of which fewer than 20 are being acquired by children and are thus potentially sustainable. Over 70% of contemporary Native American languages face imminent extinction. A revitalization movement ultimately led to the Native American Languages Act of 1992, calling for federal policy to support the cultural vitality of Native American languages and authorizing funds for their maintenance.

American Creoles

New creole languages have developed indigenously in South Carolina, Hawaii, and Louisiana. In South Carolina, a creole called Gullah or Geechee (Sea Island Creole English) began to develop in 1715 when importation of African slaves, speaking different African languages natively, increased sharply in that area. Grammatical features of the variety include: pronouns such as *ee*, *um*, *shum*, *una*; *duh* or *does be* for habitual marking; *done* to mark completed actions; null copula, null possessive, and null simple past tense. Gullah has declined in recent decades, surviving in a few coastal enclaves. As it is relegated to the home, children may speak it natively but rapidly become bilingual.

Hawaiian Creole (Hawai'i Creole English), sometimes referred to as Pidgin, began to emerge between 1790 and 1820 through contact between native

Hawaiians and Europeans; this development preceded a rise in Chinese, Portuguese, and Japanese arrivals between 1860–1900, followed by further influence from Filipino (Tagalog) and American English. These waves of contact resulted in a heterogeneous developmental process, particularly via informal and covert interaction among young speakers having English forcibly imposed on them in schools. Hawaiian Creole is characterized, among other things, by innovations in syntax (e.g., aspect marking: *stei* for progressive, *wen* for past) and in the lexicon, e.g., *pau* ‘finished’ (Hawaiian), *obake* ‘ghost’ (Japanese). Despite controversy over its societal and institutional role – only English and Hawaiian are official state languages – Hawaiian Creole is spoken and positively valued by a substantial community.

Louisiana Creole (Louisiana Creole French) is sometimes described as originally one of three French-based languages in Louisiana, alongside Cajun French (French, Cajun), brought by Acadians expelled from Nova Scotia in the 18th century, and Colonial French, an extinct variety once used by French colonizers. An alternative view treats the language situation as comprising a continuum ranging from more French to more Creole usage. The Creole arose out of contact between African slaves and French colonizers during the period of 1699 and 1750; today, due to the greater social status of English and Standard French, all Louisiana Creole speakers speak another language outside their private domains.

American Sign Language

American Sign Language (ASL) is a natural, visual-spatial language not based on American English. In 1817, the first American school for the Deaf was established, and the resulting convergence of several varieties gave rise to an expanded contact variety. By the late 19th century, an oralist movement led to the banning of signing, a situation that persisted until the 1970s. ASL use nevertheless continued throughout, sometimes covertly, and ASL is now used by 0.5–2 million people, with considerable regional and social variation.

Minority Immigrant Languages

Commonly spoken immigrant languages in the United States other than English and Spanish are listed in **Figure 2**, which shows immigration-driven reversals in language use during the 1990s: a dramatic increase in the use of Russian (192%), Vietnamese (99%), Arabic (74%), Chinese (62%) and Spanish (57%) contrasts with the decline in the use of several European languages.

European languages have been replenished by immigration since the earliest arrivals in the 15th century. The first large-scale migration of unskilled Asian laborers occurred in the mid-19th century. Chinese, Japanese, and Korean enclaves formed, while South Asian and Filipino immigrants, fewer in number of largely male, did not form self-sufficient communities as early. The second wave of Asian immigration, when quotas were extended after 1965, included refugees from Cambodia, Vietnam, and Laos as well as descendants of earlier immigrants, often more educated and economically secure than their predecessors. The majority of early Arab American immigrants were Christian; subsequent to the 1950s, there has been a rise in Muslim Arab immigration, although this group remains a minority. The major varieties of Arabic represented are Lebanese and Syrian (Arabic, North Levantine Spoken), Palestinian (Arabic, South Levantine Spoken), Egyptian (Arabic, Egyptian Spoken), and Iraqi (Arabic, Mesopotamian Spoken).

Among speakers of minority immigrant languages, or 'heritage languages,' fluency declines sharply across generations, transitioning to monolingualism within two to three generations. In particular, attrition of fluency in selected registers, shift from balanced to asymmetrical bilingualism, and decline in biliteracy across generations is widespread, largely due to institutionalized monolingualism in schools. López's (1982) findings, shown in **Figure 4**, reflect a close correspondence between the nativeness of a generation in the United States and its tendency to be English-dominant. Nevertheless, language loyalty tends to be strong across generations; in particular, **Figure 4** shows a lower rate of loss of Spanish among Mexican Americans as compared to some Asian languages. Language schools, ethnically-defined neighborhoods, and religious and cultural associations serve to maintain languages among first and second generation immigrants; third generation immigrants are generally English speakers but often show renewed, albeit often nonnative, interest in their heritage languages.

Bibliography

- Baker C & Jones S (1998). *The encyclopedia of bilingual education and bilingualism*. Clevedon: Multilingual Matters.
- Baugh J (2000). *Beyond Ebonics: linguistic pride and racial prejudice*. New York: Oxford University Press.

- Carver C M (1987). *American regional dialects. a word geography*. Ann Arbor, MI: University of Michigan Press.
- Crawford J (1999). *Bilingual education: history, politics, theory, and practice*. Los Angeles: Bilingual Services, Inc.
- Ferguson C & Heath S B (eds.) (1981). *Language in the USA*. Cambridge: Cambridge University Press.
- Finegan E & Rickford J (eds.) (2004). *Language in the USA: themes for the twenty-first century*. Cambridge: Cambridge University Press.
- Fishman J (1967). *Language loyalty in the United States*. The Hague: Mouton.
- Graddol D, Leith D & Swann J (1996). *English: history, diversity, and change*. London: Routledge.
- Green L J (2002). *African American English: a linguistic introduction*. Cambridge: Cambridge University Press.
- Kurath H (1949). *Word geography of the Eastern United States*. Ann Arbor, MI: University of Michigan Press.
- Lane H, Hoffmeister R & Bahan B (1996). *A journey into the DEAF-WORLD*. San Diego, CA: DawnSign Press.
- Labov W (1991). 'The three dialects of English.' In Eckert P (ed.) *New ways of analyzing sound change*. Orlando: Academic Press. 1-44.
- Lippi-Green R (1997). *English with an accent: language, ideology, and discrimination in the United States*. New York: Routledge.
- López D (1982). *Language maintenance and shift in the United States today* (4 vols). Los Alamitos, CA: National Center for Bilingual Research.
- McCarty T L & Zepeda O (eds.) (1998). *Indigenous language use and change in the Americas*. Special issue of the *International Journal of the Sociology of Language* (132).
- Mufwene S, Rickford J, Bailey G & Baugh J (eds.) (1998). *African American English: structure, history, and use*. New York: Routledge.
- Preston D (ed.) (1993). *American dialect research*. Philadelphia: John Benjamins.
- Roberts S J (2000). 'Nativization and the genesis of Hawaiian Creole.' In McWhorter (ed.) *Language change and language contact in pidgins and creoles*. Amsterdam and Philadelphia: John Benjamins. 257-300.
- Silva-Corvalán C (1994). *Language contact and change: Spanish in Los Angeles*. New York: Oxford University Press.
- U.S. Bureau of the Census (2003). 'Language use and English-Speaking Ability: 2000.' C2KBR-29. Washington, D.C.: U.S. Census Bureau.
- Valdman A (1997). *French and Creole in Louisiana*. New York: Plenum.
- Veltman C (1983). *Language shift in the United States*. Amsterdam, the Netherlands: Mouton Publishers.
- Wolfram W & Schilling-Estes N (1998). *American English: dialects and variation*. Malden, MA: Blackwell.
- Zentella A C (1997). *Growing up bilingual: Puerto Rican children in New York*. Malden, MA: Blackwell.

Uralic Languages

A Marcantonio, University of Rome 'La Sapienza,'
Rome, Italy

© 2006 Elsevier Ltd. All rights reserved.

The 'Uralic' languages derive their name from the Ural Mountains, the assumed homeland of the hypothetical proto-Uralic population that, according to the conventional theory, spanned out into Hungary and across a wide portion of the northern Eurasian area, from Norway to Western Siberia (see **Figure 1**).

Distribution

Of the 22 million speakers of Uralic languages, about 2 million are minority speakers in Russia. The total number of speakers is decreasing; some languages are endangered and others are now extinct. The Uralic language family can be divided into eight language subgroups:

1. Saami (formerly Lapp; 34 000 speakers); about 10 dialectal varieties are spoken in the region between Sweden and the Kola Peninsula in Russia.
2. Finnic (formerly Balto-Finnic), comprising Votic (about 50 speakers, Russia), Ingrian (400 speakers, Russia), Karelian (40 000 speakers, Finland and Russia), Lude (5000 speakers, Russia), Olonetsian (30 000 speakers, Finland and Russia), Veps (6000 speakers, Russia), Livonian (about 10 speakers, Latvia), Finnish (also called Suomi; about 5 500 000 speakers), and Estonian (about 1 000 000 speakers), including the Estonian ethnic/dialectal variety Võru-Seto (50 000 speakers) in Estonia and Russia.
3. Mordvin (Mordva; 615 000 speakers, Russia), comprising two ethnic/dialectal varieties, Erzya (about 67%) and Moksha (about 33%).
4. Mari (formerly Cheremis; 488 000 speakers, Russia), comprising two dialectal varieties, Hill (Western) Mari (about 10%) and Meadow (Eastern) Mari (about 90%).
5. Permic, or Permian (Russia), comprising Udmurt (formerly Votyak; 464 000 speakers) and Komi, consisting of three ethnic/dialectal varieties, Komi-Zyrian (217 000 speakers), Komi-Permyak

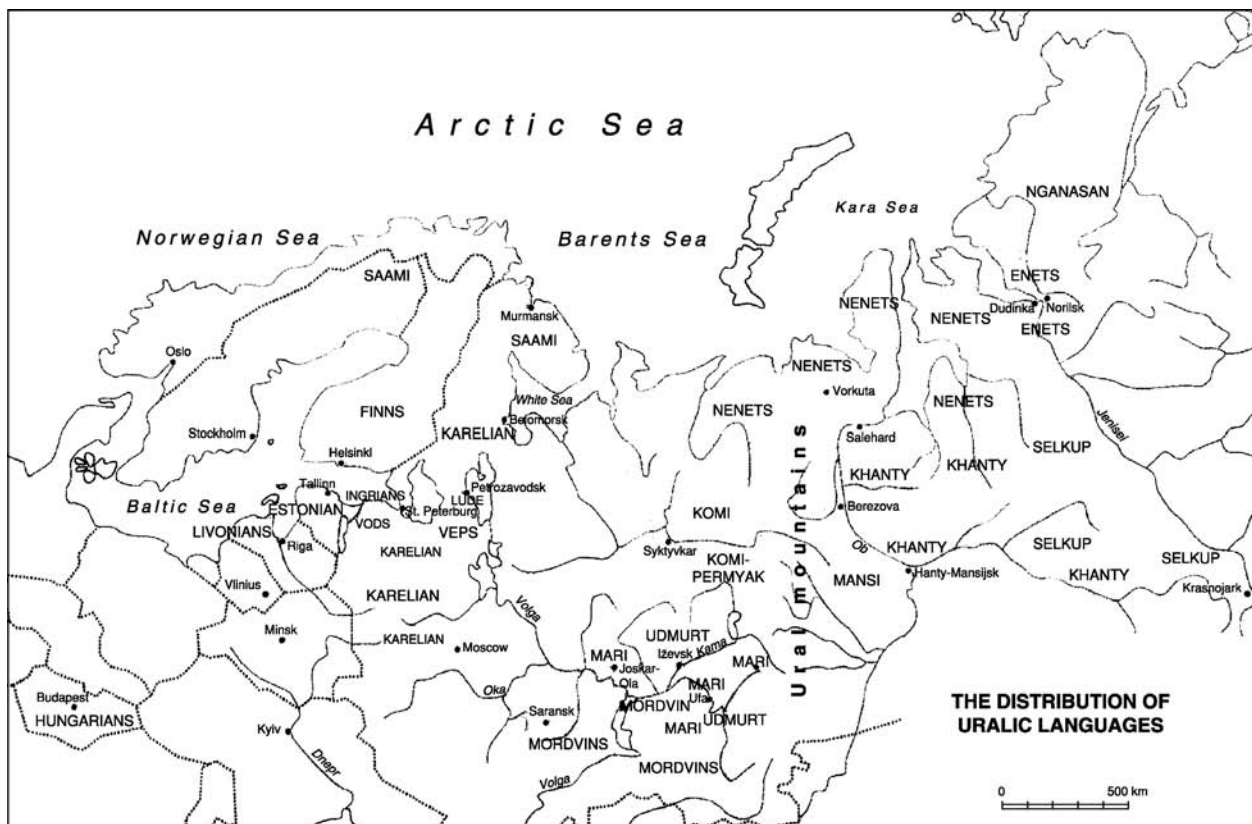


Figure 1 The Uralic languages are spoken by 22 million people. The majority consist of the Finns, Hungarians, and Estonians, living in their nation-states; some 2 million speakers are among the ethnic minorities of Russia. Reproduced from Suihkonen P (2000), *Ugriculture 2000: contemporary art of the Fenno-Ugric peoples*. Helsinki: Gallen-Kellela Museum.

- (94 000 speakers), and Yaz'va-Komi (about 200 speakers).
6. Ob-Ugric (Ob-Ugrian), comprising Mansi (formerly Vogul; 3000 speakers) and Khanty (formerly Ostyak; 14 000 speakers), scattered along the Ob' and lower-Irtysh rivers and tributaries.
 7. Hungarian (Magyar; 14 million speakers), including the ethnic/dialectal variety Csángó (100 000 speakers, Romania).
 8. Samoyed (Samoyedic), comprising seven closely related languages spoken in West Siberia, i.e., Nenets (formerly Yurak; 32 000 speakers), Enets (formerly Yenisey-Samoyed; about 200 speakers), Nganasan (formerly Tavgy; 1000 speakers), Selkup (formerly Ostyak-Samoyed; 2000 speakers), and three extinct languages, Yurats, Kamas (Kamassian), and Mator (Motor).

Hungarian and Ob-Ugric are conventionally grouped together to form the 'Ugric' subgroup, but the languages are acknowledged to be radically different in phonology, syntax, and vocabulary, and accordingly this group has not been reconstructed from the primary evidence. Several minority languages, including Veps, Mordvin, Mari, Udmurt, Komi, and Ob-Ugric, enjoy official status in their national administrative regions. Despite attempts to revitalize some endangered languages through cultural/educational/political activities and associations (e.g., 'Saami Language Nests' and 'To Save Yugra'), there remains strong pressure to assimilate into the majority languages (Suihkonen, 2002).

Phonology

Most Uralic languages display vowel harmony and consonant gradation, although there are substantial differences in implementation. These features are shared by nearby language groups, including Altaic and Yukaghir. Several Uralic languages also display quantitative vowel and consonant opposition.

Vowel Harmony

Palatovelar vowel harmony, in which the vowels of a word unit, including suffixes, enclitics, etc., are either all back or all front, is found in Finnic (not Estonian and Livonian), Mordvin, Western Mari, some Khanty and Mansi dialects, Hungarian, and Nganasan. Compare Hungarian *kert-be* 'garden-into', *kert-em-be* 'garden-my-into' and *konyhá-ba* 'kitchen-into', *konyhá-m-ba* 'kitchen-my-into'. Labial harmony occurs in Hungarian and Eastern Mari.

Consonant Gradation

Abondolo (1994: 4855) found that most Finnic and Saami languages/dialects display "alternation of

strong vs. weak consonant(ism) word-medially in open vs. closed syllables." For example, in comparing Finnish *kirkko* 'church' vs. *kirko-ssa* 'church-INESS' (INESS = inessive) and *papu* 'bean' vs. *pavu-t* 'bean-PL', the first sound of each pair, the strong grade, appears word medially in an open syllable, whereas the second sound, the weak grade, appears word medially in a syllable closed by a suffix. The Samoyed languages display a different, less homogeneous type of gradation. For example, Nganasan presents a complex co-occurrence of various mechanisms, including glottal stop alternation, truncation, syllabic and rhythmic gradation, vowel harmony, and accommodation. In some languages, within specific contexts and/or stems, the original phonetic conditioning factor for gradation has been eroded by subsequent changes; therefore, several inflectional forms can now be distinguished through grade alternation only ('fusion'). Compare the nominative (NOM), genitive (GEN), and partitive (PARTIT) in Finnish *jalka-Ø* 'foot-NOM', *jala-n* 'foot-GEN', and *jalka-a* 'foot-PARTIT' with correspondent Estonian *jala-Ø* (genitive, weak grade) and *jalga-Ø* (partitive, strong grade), in which the alternation is no longer productive.

Vocalism

The smallest vowel inventory (five vowels) is found in Erzya Mordvin; the richest inventory is found in Vakh Khanty, which has 11 full and 2 reduced, front and back (round and unround) vowels. There are diphthongs in Finnic, Saami, some dialects of Mansi, and Nganasan. In several languages, some vowels occur less frequently when not in the first syllable. Most languages (not Erzya Mordvin and most of Permian) present (some sort of) quantitative vowel opposition between two (e.g., Finnish) or three (e.g., Estonian) vowel lengths, to denote different meanings.

Consonantism

Consonantism varies considerably. Finnish has one of the smallest inventories, with 11 consonants, the obstruents being limited to the unvoiced *p*, *t*, *k*. Eastern Enontekiö (North Saami dialect) has 31 consonants; the total inventory includes voiced stops, unvoiced nasals, fricatives, affricates, palatal (or palatalized alveolar/dental) series, glides, and laryngeal and glottal stops. Several languages present quantitative consonant opposition between two-way (e.g., Finnish) or three-way (e.g., Estonian and partly Saami) opposition, to denote different meanings. Unlike the other Uralic languages, Hungarian, Permian, and (to a lesser extent) Saami display opposition of voice – for example, voiced *b* and unvoiced *p* denote different meanings.

by any numeral, and expression of subordination through nominalized/nonfinite verbal phrases (Finnish and Hungarian have recently developed subordination through conjunctions).

Basic Word Order

Saami, Finnish, Estonian, Komi, and Hungarian present as subject-verb-object (SVO); Udmurt, Ob-Ugric, and Samoyed present as subject-object-verb (SOV); Mari has a flexible order. Pragmatic/logic/stylistic functions usually play a role in determining word order.

Main Verb Phrases

Verbal phrases may be elaborated in various ways. Ob-Ugric has a passive (personal) voice – e.g., the agent is marked in Khanty by locative. Finnish uses an impersonal passive, and the agent is unspecified. In some languages, extra conjugations encode information about the object, such as number, definiteness, topicality, and referentiality. Hungarian adds one objective/definite (DEF) conjugation to the normal subjective/indefinite (INDEF) conjugation (ACC, accusative):

- (5) olvaso-k
read-1SING.INDEF
'I read (something)'.
- (6) olvaso-m
read-1SING.DEF
'I read it'.
- (7) olvaso-m a könyv-et
read-1SING.DEF the book-ACC
'I read the book'.

Ob-Ugric adds three objective conjugations, for singular, dual, and plural objects. Nenets, Enets, and Nganasan have five conjugations: one subjective, three objective (as Ob-Ugric), and one objectless/reflexive. The markers differ.

Subordinate Sentences

There are several types of nonfinite (participial, infinitival, and gerundive) verbal phrases. Typically, the verb takes the relevant nonfinite morpheme, and then may be inflected with enclitics, and case, number, possessive, and passive suffixes. Compare Finnish, in which *-ä* (~ *-a*) and *-ma* (~ *-mä*) are infinitive morphemes (TRANSLV, translativ; EL, elative):

- (8) syö-mme elä-ä-kse-mme
eat-1PL live-INF-TRANSLV-1PL
'We eat to live'.
- (9) Pekka on koto-na leikki-mä-ssä
Pekka is home-at play-INF-INESS
'Pekka is at home playing'.

- (10) Pekka tuli puutarha-sta ä-stä
Pekka came garden-from play-INF-EL
'Pekka came from the garden from playing/
where he was playing'.

Objects

Marking of the direct object is varied and complex, often depending on pragmatic/aspectual factors (as in Examples (3) and (4)), or on the type of sentence the object is in – for example, Hungarian has *-t*, Khanty and Sosva Mansi have *-Ø*, Eastern Mari and some Samoyed languages have *-m*, Finnish has *-n* or partitive for singular and *-Ø* or partitive for plural objects, and Udmurt has *-Ø* for indefinite and accusative for definite objects. Number agreement occurs between subject and predicate. Within the noun phrase, agreement in number and case suffixes occurs in some languages and to various degrees of completeness, being fully developed in Finnish.

Uralic Languages as a Family

The results of recent archaeological, genetic, and anthropological research are inconsistent with the predictions of the Uralic theory, and the significance of the linguistic evidence on which the conventional theory is based has been called into question: for example, there is no reconstruction of the key Ugric node based on the primary evidence, and the common linguistic tendencies appear to be shared with other language groups, such as Altaic. Alternative models have been proposed (see Künnäp, 2000; Marác, 2004; Marcantonio, 2002; Wiik, 2002).

Bibliography

- Abondolo D (1994). 'Uralic languages.' In Asher R E (ed.) *The encyclopedia of language and linguistics*. Oxford: Pergamon Press. 4855–4858.
- Abondolo D (ed.) (1998). *The Uralic languages*. London: Routledge.
- Collinder B (1965). *An introduction to the Uralic languages*. Berkeley: University of California Press.
- Comrie B (1981). *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Hajdú P (1975). *Finno-Ugric languages and people*. London: Deutsch.
- Hajdú P & Domokos P (1987). *Die uralischen Sprachen und Literaturen*. Hamburg: Buske.
- Korhonen M (1996). *Typological and historical studies in language. A memorial volume*. 223. Helsinki: La Société Finno-Ougrienne.
- Künnäp A (2000). *Contact-induced perspectives in Uralic linguistics*. München: Lincom Europa.
- Marác L (2004). 'De oorsprong van de Hongaarse taal.' In van Heerikhuizen A *et al.* (eds.) *Het Babylonische*

- Europa*. Amsterdam: Amsterdam University Press. 81–96.
- Marcantonio A (2002). *The Uralic language family. Facts, myths and statistics*. Oxford: Blackwell.
- Sinor D (ed.) (1988). *The Uralic languages. Description, history and foreign influences*. Leiden: Brill.
- Suihkonen P (2002). 'The Uralic languages.' *Fennia* 180, 165–176.
- Wiik K (2002). *Eurooppalaisten juuret*. Jyväskylä: Atena.

Relevant Websites

- <http://www.helsinki.fi> – Helsinki home page, with links to a classification by Tapani Salminen of the Uralic (Finno-Ugric) languages.
- <http://www.suri.ee> – Website on the history of the Finno-Ugric peoples.

Urdu

H Dua, Central Institute of Indian Languages, Mysore, India

© 2006 Elsevier Ltd. All rights reserved.

Urdu is the literary, cultural, and religious language of Muslims in India, Pakistan, Bangladesh, and other parts of the world including the United States, the United Kingdom, Germany, and Sweden. The number of Urdu speakers in census data may be under- or overestimated for social and political reasons. However, it is estimated that Urdu is spoken by 54 million worldwide, out of which 43 million speakers are found in India. In addition to being the national language of Pakistan, Urdu is one of the Schedule VIII languages of the Indian democracy, the state official language of Jammu and Kashmir, and the second official language of UP, Bihar, and Andhra Pradesh in India. It is recognized that Urdu, Hindi, and Hindustani share a common grammatical system. Urdu in its colloquial form may therefore be considered the lingua franca of one of the largest speech communities in the world. Urdu is regarded as a pluricentric language that shows different linguistic features.

Origin and Development

Historically, Urdu has developed in a language contact situation over a long period from 1100 A.D. or earlier. After the Muslim invasion of India, it emerged as a speech variety in communication among Muslim rulers, traders, mystics, and the local population. The early form of Urdu developed out of the literary language Sauraseni Apabrahṃsa, which was in a state of transition and developing as a New Indo-Aryan language. It had a wide dialect base that included Braj Bhasha, Haryanvi or Bangaru, eastern Panjabi, and other dialects spoken in the region surrounding Delhi. Khari Boli was present as one of the elements in the

formative period of Urdu and it gradually became stronger with its development. By 1800, Khari Boli could be considered as the basic source of Urdu. During the period of development, from 1100 to 1800 A.D., Urdu was known by several different names, including Hindwi, Dehalvi, Hindustani, Zaban-e-Urdu, Dakhini or Old Urdu, and Rekhta. The first use of the language name Urdu was made in a couplet in 1776 by the poet Mashafi (1750–1824). However, the use of Urdu, referring to camp, court, or city (Zaban-e-Urdu or Zaban-e-Urdu-e-Shahi or Zaban-e-Urdu-e-Mualla), had been in use from 1560.

Specimens of Hindwi in the early formative period are found scattered in the Nath Panthi literature, early Sufis of North India, Amir Khusro, Nanak, Kabir, Baba Farid, and other poets. Amir Khusro (1236–1324) shows a distinct earlier form of Urdu, or Hindwi as he calls it. However, there is no evidence that the language was in continuous use from 1200 to 1650 except *Bikat Kahani* by Afzal, which appeared 300 years after Amir Khusro's writings. It is therefore not possible to reconstruct a continuous history of the development of Urdu (Chatterji, 1960; Khan, 1958). Insha Allah Khan Insha's *Darya-e-Latafat* ('The river of elegance,' 1807) presents an early linguistic study of the dialects of Delhi and Lucknow.

The emergent variety Hindwi traveled in the south with the Muslim rulers of the Delhi Sultanate (1211–1504) along with the Muslim armies, traders, Sufis, preachers, and other people. It flourished as a literary language in the Deccan kingdoms of Golkunda and Bijapur. It was popularly known as Dakhini or Hindwi or Dehalvi. Dakhini has been claimed as Dakhini Hindi, Dakhini Urdu, or Old Urdu. It shows some linguistic features that are characteristic of its contact with local languages of the South. However, the origin and development of Dakhini has been traced to Haryanvi, Panjabi, Braj Bhasha, and Khari Boli.

During the Mughal period, Persian was the official language of the court. The elite and noblemen spoke and wrote Persian. The Mughal emperors from Akbar onward spoke an early form of Hindustani at home (Chatterji, 1960). Both the Hindus and the Mughal rulers accepted Braj Bhasha as a literary language. Akbar's courtier, Khan Khanan Rahim, wrote in Braj Bhasha, and even Akbar attempted to write some verses in it. Hindustani or Khari Boli did not develop as a literary language in the north until Wali from Aurangabad arrived in Delhi at the end of the 17th century. Wali demonstrated that Hindustani with a scattering of Persian words could be used to write great poetry. The language used by Wali is known as Rekhta, which means 'scattered,' and implies that Hindustani or Urdu had not been 'Persianized.' The Delhi school of poetry came into existence around Wali during 1700–1720.

By the end of the 18th century and the beginning of the 19th, Urdu, in its modern form, had taken deep roots. Several factors contributed to the emergence of Urdu in its distinct modern form. First, the Dakhini literature was written in the Perso-Arabic script, which had "fixed the orientation of the language" (Chatterji, 1960). Wali and subsequent poets and writers readily adopted the Perso-Arabic script. It became a symbol of linguistic and cultural identity. Second, conscious efforts were made by stalwarts such as Khan Arzu (1689–1756), Shah Hatim (1699–1781), and Mazhar Janejani (1700–1781) to weed out the Braj Bhasha or indigenous words from Rekhta and incorporate Arabic and Persian words in it during the middle of the 18th century. The extreme Persianization of Urdu became characteristic of the Lucknow school of poetry, whereas the Delhi school developed its own standard form of Urdu. Third, the use of subjunctive constructions, the continuous tenses with 'raha,' the ergative construction with the postposition 'ne,' and the formation of the present tense with imperfect participles became stable and characteristic. Finally, prose began to be written in the emergent Khari Boli by the end of the 18th century. The establishment of Fort William College at the beginning of the 19th century encouraged the development of two styles of prose that paved the way for the emergence of Hindi and Urdu as distinct standard varieties. The two important earliest works in Urdu prose are the *Bagh-o-Babar* of Mir Amman (1804) and the *Khairat Afroz* of Hafizuddin Ahmed (1803–1815).

Urdu Language: Identity and Conflict

The 19th century may be considered to be the century of consolidation, expansion, and growth of Urdu language identity and literature, on the one hand, and

the spread of Urdu and sociopolitical mobilization, on the other. Several mutually interactive forces played a catalytic role in this regard. A synoptic view of some of these factors reflects this. First, after the early prose produced at Fort William College, Urdu literature developed rapidly. All genres of literature, including novel, short story, drama, different forms of prose, and journalistic forms developed and made distinctive achievements. Urdu poets throughout the 19th century flourished, and Urdu entered the modern period with Hali (1837–1914) and Akbar Allahabadi (1846–1921) as well as many others. Muhammad Hussain Azad (1830–1910), in *Ab-e hayat* (1880), provided the first systematic account of the achievements of Urdu poetry, constructing a literary history, a canon, and the theory of poetry. Second, the establishment of several educational institutions, including Delhi College, Anjuman-e-Punjab, and Mohammedan Anglo-Oriental College, played multiple roles in enriching Urdu literature with translations from English as well as original writings in different disciplines. This trend spread the use of Urdu language and literature and contributed to the development of linguistic, literary, and cultural identity. Third, Urdu language and literature gained in momentum when it replaced Persian in 1837. It was used as an official court language along with English in the British-ruled provinces in North India. It gave rise to what is popularly known as the Hindi movement. Between 1868 and 1900, the Hindus of the northwestern provinces fought against Urdu through pamphlets and memoranda. They argued that the Perso-Arabic script was alien to India, that it was unintelligible to common people, and that Hindi written in the Devanagari should be made an official language. As a result of the agitation, in 1881 Hindi replaced Urdu in Devanagari script as the official language of the neighboring province of Bihar. This paved the way for the hardening of cultural-communal attitudes among the speakers of Urdu and Hindi, the divergence of Hindi and Urdu, and the formation of different linguistic identities. This can be seen in the exclusion of Hindu poets and the Hindu community in constructing the history of Urdu literature, on the one hand, and the switching of Hindu writers from Urdu to Hindi on the other (Faruqi, 2001). Prem Chand's switch from Urdu to Hindi was not merely an individual, personal choice but also was intricately involved with interrelated linguistic, political, and economic developments. This process reached its culmination with the complete identification of Urdu with Muslims in the second quarter of the 20th century.

The conflict between Urdu and Hindi was aggravated by the end of the 19th century and in the second

quarter of the 20th century. Two factors played a significant role in this process. First, this period saw the development of voluntary language associations such as Nagari Pracharini Sabha, formed in 1893, Hindi Sahitya Sammelan, founded in 1910, and Anjuman-Taraqqi-e-Urdu, formed in 1903. These associations promoted the cause of Hindi and Urdu, divided the loyalties of Hindi-Urdu speakers and writers, strengthened the linguistic divisions, and consolidated separate identities. Second, the Hindi-Urdu conflict and identities were reinforced by the development of both Hindu and Muslim revivalism and communal antagonism in the context of the Western culture, on the one hand, and the growth of the independence movement, on the other. As a consequence, the political mobilization of the masses contributed to the congruence of symbols of linguistic, cultural, and linguistic identities with the process of nationalism and nation formation. Das Gupta (1971: 57) points out that the identification of nationalism, linguistic, and religious solidarity was “more integral and pervasive” in the case of Muslims as compared to that of the Hindus. Ultimately, the partition of India led to the development of Urdu language and literature in India and Pakistan along different lines. This resulted in two linguistic and literary consequences. First, both the Hindi and Urdu speakers gave up Hindustani on ideological grounds. Although Hindi speakers identified Hindustani with Urdu, the Urdu speakers considered it another form of Hindi. Second, both the Hindi and Urdu speakers lost sensitivity and ability to appreciate the literature in a language other than their own.

Linguistic Description

It is generally recognized that Hindi and Urdu share a common grammatical system. They differ mainly in their writing systems, in their lexicon borrowed from Sanskrit or Persian and Arabic resources, and the minor aspects of syntax. Thus, at the phonological level, Urdu has a subset of phonemes (f x š z ž γ q) because of Perso-Arabic words, whereas Hindi has acquired ŋ š ʃ from Sanskrit words. Kelkar (1968: 80) points out that “it is highly unlikely that H š ŋ i on the one hand, U q ? z x on the other will coexist in the same idiolect.” Similarly, Urdu has acquired some other distinctive phonological features. Khan (1978: 10–11) points out that Urdu speakers invariably break up consonant clusters in VCC structure in words of Sanskrit origin, but they pronounce the structure correctly in the Persian and Arabic words. This is partly because of cultural influence of Perso-Arabic vocabulary, and partly because of the educational background of speakers. This refers to the phenomenon of *Schwa* deletion having a wider

scope in phonological analysis. Narang and Becker (1971) show that a group of derived nouns and adjectives of Perso-Arabic origin represent an exception to the *Schwa* deletion rule. In short, the distinctive phonological features of Urdu are mainly due to Perso-Arabic words. However, it is generally not specified whether these features are characteristic of written or spoken style or both, or educated or uneducated speakers of Urdu.

The issue of lexical borrowing raises different problems at the lexical or semantic level. Borrowed words may be considered in terms of word classes such as nouns, adjectives, adverbs, compound verb formatives, and so on. For instance, Hindi and Urdu show a clear difference in compound verbs consisting of noun + verb or adjective + verb sequences such as U šūrū karnā, istēmāl karnā and H ārambh karnā and prayōg karnā. It is essential to highlight some important issues that are not discussed in the analysis of lexical differences between Hindi and Urdu. First, the studies of distinctive lexicon are based on restricted data as evident from Mobbs (1981) and van Olphen (1989). The implications of the nature and scope of lexical differences between Urdu and Hindi can be understood only on the basis of a large representative sample of both spoken and written varieties belonging to different forms of literature. The corpus of three million words, each in Urdu and Hindi, available with the Central Institute of Indian Languages, Mysore, offers challenging opportunities for a wide range of linguistic studies. Second, it is necessary to recognize that the choice of a word of Persian or Arabic origin does not necessarily imply a choice in favor of Urdu. Similarly, the use of words of Sanskrit origin does not imply Hindi. In other words, both Perso-Arabic and Sanskrit words may have been assimilated and become part of the primary system of Hindustani and thus constitute an integral feature of both Hindi and Urdu. Finally, it is essential to move beyond individual lexical items and bring out the implications of borrowed words in collocations and in reflecting different cultural meanings, values, and history. In other words, it is essential to explore to what extent different sets of Perso-Arabic or Sanskrit vocabulary individually as well as in different collocations contribute to the construction of different conceptualization of entities, events, and situations, and different worldview, at the semantic level. Prem Chand's switch from Urdu to Hindi clearly shows how he found Sanskrit vocabulary congenial to the themes of his works and the sociocultural worldview related with them (Trivedi, 1989).

The borrowing of Perso-Arabic words in Urdu creates characteristic linguistic features at the grammatical level. This can be seen in a number of

word-forming suffixes and prefixes. The process of compound formation in Persian has contributed to productivity of compounds in Urdu. Similarly, the process of word formation in Urdu depends a great deal on Arabic resources. This is particularly evident in the various derived verbs with their associated participles and verbal nouns along with their word forming affixes and vowel patterns added to the root. In other words, the productive process of word formation characteristic of Urdu at the grammatical level shows a deep impact of Perso-Arabic resources. Similarly, the distinctive grammatical features of Urdu can be seen in the use of some prepositions, negative particles, formation of noun duals, or plurals in the case of some nouns that are a result of Perso-Arabic influence. The grammatical analysis of Urdu cannot ignore these linguistic devices, as they are extremely productive and provide a distinctive character. However, a number of issues remain to be explored. First, it is essential to explore how deeply these linguistic devices have influenced the structure of Urdu. It will be useful to study whether these features are particularly typical of literary or administrative language, or newspaper texts, or they are also found in everyday Urdu use. Second, it would be relevant to explore the extent to which these linguistic devices support the process of divergence of Urdu from the colloquial Hindustani grammatical system. In this respect, van Olphen (1989) points out that "it is convergence that threatens Urdu in India." By contrast, Hasnain (1995) shows in a small-scale empirical study of Urdu used in mass media and education that innovations in language based on Perso-Arabic resources of word formation have implications for comprehension or intelligibility of language use. In short, whereas language innovations based on Perso-Arabic resources may contribute to divergence of Urdu from the Hindustani grammatical system, the pressure of comprehensibility may check the trend of divergence. The consequences of the dynamics of convergence and divergence will become clear only in the long run.

Codification and Standardization

Language planning agencies and organizations played a significant role in the development and standardization of Urdu. Anjuman Taraqq-e-Hind, established in 1903, has been in the forefront in the development and promotion of Urdu. After the partition of India, the reorganized organization was less militant and more concerned with the promotion and popularization of Urdu among the people. It has 10 branches in different states and eminent leaders, such as Kazi Abdul Gaffar and Zakir Hussain, have played an

important role in its growth. It has made a significant contribution for the recognition of Urdu as a second official language in UP and Bihar and for the extension of its use in schools, colleges, and in radio communication. It has been engaged in the organization of celebration of Ghalib (1797–1869), Iqbal (1878–1938), and Prem Chand (1880–1936) days to popularize Urdu and Urdu conventions involving educational, literary, social, cultural, and political societies (Brass, 1975). Another organization, the Deeni Talimi Council, has focused its attention on the contents of textbooks. It works for the preservation of Muslim cultural values and basic tenets of Islam. Jamia Milia Islamia, established in 1920, has become one of the important educational and academic institutions concerned with Urdu education and academic research. The University Grants Commission recognizes it as a 'central university.' It not only gained prestige and respectability in Urdu education and studies but also played a constructive role in support of Urdu by influencing the language policy of the Union Government (Das Gupta, 1970).

In addition to the nongovernmental organizations, the central and state governments have made a significant contribution to the development and standardization of Urdu. The Bureau for the Promotion of Urdu, established by the Government of India in 1969, has done extensive work on the codification and standardization of Urdu. It has produced 100 000 Urdu technical terms for various disciplines of natural sciences, social sciences, and art, published more than 600 books on academic subjects, and compiled Urdu-Urdu and English-Urdu dictionaries, and Urdu encyclopedias. In UP, Bihar, Madhya Pradesh, Maharashtra, and other states, Urdu academies have been working on translation of books from English, publication of standard literary and scholarly works, university level textbooks, and the promotion of Urdu through seminars and conferences. Similar work on the codification and standardization of Urdu has been going on in Pakistan. The evaluation of the extensive work on development, codification, and standardization of Urdu needs to be studied, focusing on the impact of this on language change and development of pluricentric norms in the two countries. This is also relevant from the point of view of divergence of standard Urdu from the colloquial norm and its implications for comprehension by educated speakers of Urdu.

Although the codification and standardization work by both the government and nongovernmental organizations is essential and significant, it is also important to recognize the contribution of the individuals as creative writers, researchers, scholars, educationalists, linguists, and teachers who play a critical

role in the stabilization and cultivation of the standard language. It is not possible to mention all the names of Urdu specialists who have made a substantive contribution to research and development of Urdu. It may, however, be mentioned that several eminent scholars and researchers on Urdu have been recognized for their seminal contribution in various fields of studies on Urdu including Urdu script and spelling reform, lexicography, standardization of pronunciation and vocabulary, historiography of Urdu language and literature, and linguistic analysis and description.

Urdu Literature

Literature has been one of the most significant sources of language development and standardization in the case of many developed languages of the world. The history of Urdu shows parallel development of both literature as well as language. Just as Urdu language was formed in communication and social interaction between two cultures in the situation of language contact, Urdu literature shows fusion of two literary and cultural traditions. The Perso–Arabic elements in Urdu language do not merely constitute a superimposed structure but also form an integral aspect of language identity and its literary tradition. They have a rich semantic potential expressive of Islamic tradition and cultural worldview. Similarly, Urdu literature shows a synthesis between Islamic and Indian cultural traditions at literary, aesthetic, and philosophical levels. Narang (1991) maintains that although Urdu literature has been deeply influenced by Persian literature and rich Iranian and Islamic tradition, it has imbibed Indian cultural influences and has emerged as an expression of the composite culture of India. This is evident from the development of various forms and genres of literature during the last 300 years.

Medieval Urdu poetry shows profound influence of Persian literary tradition in its various forms, imagery, and figures of speech, as well as themes and background. The *ghazal* in the medieval poetry has “no local color,” lacks personal touches, and appears to have largely a “museum” quality (Sadiq, 1984). However, in the process of development of Urdu literature over the next two centuries, *ghazal* grew beyond erotic themes. Sadiq (1984: 19) points out that “nothing seems to be alien to its genius and it has readily accommodated ethics, metaphysics, philosophy, mysticism, satire, politics, side by side love, which still continues to be its favourite theme.” The semiotic analysis of *ghazals* of Ghalib, Iqbal, Faiz, and Firaq Gorakhpuri (1896–1982) brings out the rich potential of the genre of *ghazal*. The same is

true of other genres such as *Masnavi*, *Marsiya*, *Qasida*, and so on. *Masnavis* of Mir Hasan (1727–1786) are soaked with Indian imagery (Sadiq, 1984: 16). Srivastava (1992) shows that *Masnawi* as a poetic form has assimilated in its content Puranic legends, Indian folktales, semihistorical events of Indian soil, and so on in the process of its endogenous growth. The popular love lyric *Qawwali* was not only exclusively developed in India but also became an integral part of the secular North Indian music gaining in popularity, as did its indigenous counterparts such as Hindu Kirtan or Bhajan (Narang, 1991). Mushaira (poetic symposia) has become a popular literary convention in India, Pakistan, and other parts of the world.

The modern Urdu literature has many great achievements to its credit. The individual achievements of great poets, writers, and men of letters are difficult to enumerate. However, it would be adequate to mention a few points that are characteristic of the vitality of Urdu literature. First, the development of Urdu literature has kept pace with the trends and tendencies of the time and produced poets and writers belonging to different traditions, movements, and ideologies. Similar to literary traditions in other major Indian languages, it represents a great deal of involvement, sensitivity, and an awareness of contemporary social reality. For instance, the progressive story writers in Urdu, Rajendra Singh Bedi, Manto (1913–1955), Krishan Chander (1914–1977), and Ismat Chughtai (1915–1991), give expression to economic inequality, social exploitation, and male chauvinism as do their counterparts in Hindi. Quarratul-ain-Haider (b. 1928) and Ismat Chughtai in Urdu focus on Indian women and their consciousness as do Manu Bhandari and Krishna Sobti in Hindi. The Sahitya Akademi award to Ismat Chughtai and the Jnanpeth award to Quarratul-ain-Haider have gained recognition for Urdu literature at the national level.

Second, Urdu literature is not merely restricted to poetry or fiction. It encompasses a wide range of literary criticism, folk literature, children’s literature, and scientific literature. In terms of total literary output, Urdu does not lag behind several major Indian languages.

Finally, it is worth emphasizing that several Urdu poets and writers have carried forward the tradition of the synthesis of the Islamic and the Indian cultures. In this context, Salahuddin Pervez has achieved a great distinction in his novel *Identity Card* for giving expression to the spirit of Islamic thought and its interaction with the Indian spiritual–cultural system and transforming it into a powerful universal humanistic Indo-Islamic ethos. There is a distinct progress

in sincere appreciation and creative assimilation of Buddhist thought and its cultural tradition. Both poets and fiction writers discover a rich potential of myths, jataka tales, and Buddhist philosophy. Quarratul-ain-Haider in her masterpiece *Aag Ka Darya* ('The river of fire') makes an imaginative representation of Vedic and Buddhist elements in the spiritual saga of man. Khalilur Rahman Aazmi (d. 1978), one of the pioneers of the new movement in Urdu poetry, presents Gautam as a symbol of perfection. Similarly, Yusuf Zafar, a leading figure of New Poetry in Pakistan, portrays Buddha as an embodiment of love and compassion and a landmark in the spiritual history of mankind.

In short, Urdu literature shows a genuine creative assimilation of the ancient Indian cultural tradition and philosophy in the context of the contemporary problems of mankind in modern age. The standardization and elaboration of the Urdu language shows not only its communicative dynamics and expressive potential but also the loyalty and identity of its speakers. Thus, the Urdu language and literature have gained recognition because of their vitality and achievements and spread at the international level.

Bibliography

- Bailey T G (1928). *A history of Urdu literature*. (rpt. 1979). Delhi: Longman.
- Beg M K A (1995). 'The standardization of script for Urdu.' In Hasnain I S (ed.) (1995) *Standardization and modernization*. New Delhi: Bahari Publications. 227-241.
- Brass P R (1975). *Language, religion and politics in north India*. Delhi: Vikas Publishing House Pvt. Ltd.
- Chatterji S K (1960). *Indo-Aryan and Hindi*. Calcutta: Mukhopadhyay.
- Das Gupta J (1970). *Language conflict and national development: group politics and national language policy in India*. Berkeley and Los Angeles: University of California Press.
- Das Gupta J (1971). 'Language, religion and political mobilization.' In Rubin J & Jernudd B H (eds.) (1971) *Can language be planned?* Honolulu: University Press of Hawaii.
- Dua H R (1992). 'Hindi-Urdu as a pluricentric language.' In Clyne M (ed.) *Pluricentric languages*. Berlin and New York: Mouton de Gruyter. 381-400.
- Dua H R (1995). 'Sociolinguistic processes in the standardization of Hindi-Urdu.' In Hasnain I S (ed.) *Standardization and modernization*. New Delhi: Bahari Publications. 177-196.
- Faruqi S R (2001). *Early Urdu literary culture and history*. New Delhi: Oxford University Press.
- Hasan M (2002). 'Buddhist elements in Urdu literature.' *Indian Literature* 208, 168-180.
- Hasnain I S (1995). 'Innovations in language – an experiment in comprehensibility with reference to Urdu in mass media and education.' In Hasnain I S (ed.) *Standardization and modernization*. New Delhi: Bahari Publications. 213-226.
- Kachru Y (1990). 'Hindi-Urdu.' In Comrie B (ed.) *Major languages of South Asia, the Middle East and Africa*. London: Routledge. 53-72.
- Kelkar Ashok R (1968). *Studies in Hindi-Urdu: introduction and word phonology*. Poona: Deccan College.
- Khan M H (1969). 'Urdu.' In Sebook T A (ed.) *Current trends in linguistics*, vol. V. The Hague: Mouton. 277-283.
- Khan M H (1978). 'A phonetic and phonological study of the word.' In Singh K S (ed.) *Readings in Hindi-Urdu linguistics*. Delhi: National Publishing House. 3-33.
- King C (1994). *One language, two scripts: the Hindi movement in nineteenth century north India*. Delhi: Oxford University Press.
- Mobbs M C (1981). 'Two languages or one? the significance of language names "Hindi" and "Urdu."' *Journal of Multilingual and Multicultural Development* 2(3), 203-211.
- Narang G C (1991). *Urdu language and literature: critical perspectives*. Delhi: Sterling Publishers.
- Narang G C & Becker D A (1971). 'Aspiration and nasalization in the generative phonology of Hindi-Urdu.' *Language* 47(3), 647-667.
- Pritchett F & Faruqi S R (2001). *Ab-e hayat shaping the canon of Urdu poetry*. New Delhi: Oxford University Press.
- Rai A (1984). *A house divided: the origin and development of Hindi/Hindawi*. Delhi: Oxford University Press.
- Sadiq M (1984). *A history of Urdu literature*. Delhi: Oxford University Press.
- Schmidt R L (1999). *Urdu: an essential grammar*. London and New York: Routledge.
- Schmidt R L (2003). 'Urdu.' In Cardona G & Jain D (eds.) *The Indo-Aryan languages*. London and New York: Routledge.
- Srivastava R N (1969). Review of Kelkar, Ashok R. (1968). *Studies in Hindi-Urdu: introduction and word phonology*. Poona: Deccan College. *Language* 45, 913-927.
- Srivastava R N (1992). Review of Narang, Gopi Chand (1992). *Urdu Language and literature: critical perspectives*. Delhi: Sterling Publishers.
- Trivedi H (1989). 'The Urdu Premchand: the Hindi Premchand.' In Mohan C (ed.) *Aspects of comparative literature*. New Delhi: India Publishers and Distributors.
- van Olphen H H (1989). 'Lexical convergence in Urdu and Hindi.' In *Paper presented in the International Seminar on the Common Bases of Urdu and Hindi*. Aligarh: Aligarh Muslim University.

Uto-Aztecan Languages

C Fowler, University of Nevada, Reno, Nevada

© 2006 Elsevier Ltd. All rights reserved.

Uto-Aztecan is a large family of indigenous languages whose descendants are distributed from Oregon in the north to El Salvador in the south, with the heaviest concentrations of contemporary speakers in northern and central Mexico. Today over 45 extant and extinct languages are recognized as part of the family, with some of the extant languages represented by 50 or fewer speakers and others by well over 100 000 (Campbell, 1997; Ethnologue, 2004). In times past, speakers of these languages included peoples displaying the full range of socioeconomic adaptations, from small extended families who lived by hunting and gathering to clans of small village farmers to intensive agriculturalists organized into vast empires. Today, in many communities in the United States, public education and culture change have reduced the number of speakers to dangerously low levels, and language extinction appears inevitable. For others, concerted efforts at language salvage and revitalization that are currently underway may prolong or actually reverse the decline. For several of

the more remote and robust languages of Mexico the picture is brighter and there appears to be less danger of significant language loss in the immediate future.

The languages within the Uto-Aztecan family are divided into three more or less contiguous geographic clusters across their broad range. The names for each cluster and views as to their internal relationships have changed through time and are still subject to some debate. The units are generally referred to as (1) Shoshonean or Northern Uto-Aztecan, which includes several branches and languages concentrated in the Great Basin and southern California in the United States; (2) Sonoran, which includes the languages of southern Arizona in the United States and of Sonora, Chihuahua, and Durango in northwest Mexico; and (3) Aztecan or Nahuatl, with languages widespread in central Mexico and outliers in El Salvador. The Sonoran and Aztecan languages are often subsumed under the term Southern Uto-Aztecan, either as a geographic reference to contrast them with the languages of the north or in recognition of a genetic relationship. The languages within each of the clusters and the branches with which they are affiliated are given in **Table 1**.

Table 1 Uto-Aztecan languages

Northern Uto-Aztecan

Numic:

- Western [2 languages = Mono (Monache, Owens Valley Paiute) and Northern Paiute (including Bannock)]
- Central [3 languages = Panamint (Timbisha), Shoshone (Western, Northern, Eastern, Gosiute) and Comanche]
- Southern [2 languages = Kawaiisu and Ute (Northern and Southern Ute, Southern Paiute, Chemehuevi)]

Takic:

- Serrano-Gabrielino [3 languages = Serrano (Vanyume), Kitanemuk, *Gabrielino (Fernandeño)]
- Cupan [3 languages = Cahuilla, Cupeño, Luiseño (Juaneño)]
- *Tataviam (?)

Tubatulabal

Hopi

Southern Uto-Aztecan

Tepiman:

- Upper Piman [1 language = (Pima, Tohono O'odham, Nevome)]
- Lower Piman [1 language = (Mountain, Yepachi, Yecora-Maycoba)]
- Northern Tepehuan
- Southern Tepehuan [1–3 languages = (Southern Tepehuan, Tepecano)]

Tarahumitan:

- Tarahumaran [2–7 languages = Tarahumara (Western, Northern, Central, Southern, Ariseachi, Summit) and Guarijio (Upland, Lowland)]
- Opatan [2 languages = Opata and Eudeve]
- Cahitan [1–2 languages = Yaqui and Mayo]

Corachol:

- Cora [1–2 languages (Cora, Santa Teresa Cora)]
- Huichol

Aztecan:

- *Pochutla
- General Aztec (4–28 languages/dialects = Pipil, Nahuatl (Mexicano, Aztec, Tetelcingo, Zacapoaxtla, etc.)

After Campbell, 1997; Goddard, 1996; Ethnologue, 2004; does not include all extinct* languages.

Records and studies of Uto-Aztecan languages extend back to the time of the Spanish conquest of Mexico with the compilations of Fray Bernardo de Sahagun from 1540 to 1560 on Classical Aztec or Nahuatl, as well as by other early pioneers. Documentation of most of the northern languages did not begin until exploration and colonization of the western United States in the first decades of the 19th century. As explorers, military expeditions, traders, and missionaries began to compile vocabularies of western U.S. and northern Mexican languages, the work on genetic classifications of them began in earnest. Initial compilations by Albert Gallatin in the 1830s and 1840s, Johann Carl Buschmann in the 1850s, and Albert Gatschet in the 1870s, among others, led to the inclusion in the family of most of the languages and branches recognized at present. Most controversial was the linking of Nahuatl to the Sonoran and ultimately Shoshonean languages, proposed by Bushmann and accepted by Gatschet in 1878, but rejected by John Wesley Powell in his classification of 1891. Daniel Brinton in his classification the same year accepted the linkage and is credited with naming the family, choosing a northern language (Ute) and a southern one (Aztec) to represent the unity (Goddard, 1996; Lamb, 1964). However, not until Edward Sapir (1913–1914) provided the first systematic study of sound correspondences and lexical reconstructions within the family by comparing Southern Paiute with Nahuatl was the overall relationship considered to be demonstrated. Since that time, work has concentrated on better understanding internal and external relationships for the family, and on the basic description and documentation of the individual languages (Goddard, 1996; Lamb, 1964).

A link between the Uto-Aztecan language family and the Tanoan languages of the U.S. Southwest and thus ultimately to Kiowa of the U.S. Plains was first suggested by Sapir in his 1929 macro-classification of North American Indian languages. This grouping was referred to by him as Aztec-Tanoan, and later given the position of a phylum or superstock. Benjamin Lee Whorf and George Traeger provided sound correspondences and reconstructions to show the Tanoan linkage, although they initially rejected the inclusion of Kiowa. The Kiowa-Tanoan linkage was confirmed in the late 1950s and early 1960s (Goddard, 1996: 313, 317), but the Aztec-Tanoan combination has not fared as well. Suggestions of this and yet more remote relationships for Uto-Aztecan languages, all of which appear doubtful, are reviewed by Campbell (1997).

The internal relationships of some of the Uto-Aztecan branches and sub-branches are still debated. The combinations of languages that make up the

northernmost branch, Numic, found in the Great Basin of the United States, are solid. Two languages, Hopi of northern Arizona and Tubatulabal (Tübatulabal) of southern Sierran California, are understood to be independent branches. Early extinctions and thus lack of data for some of the languages of the Takic branch of southern California make internal relationships difficult to determine with certainty, particularly the position of Garbrialiño and Tataviam (Campbell, 1997: 135). Powell and A. L. Kroeber suggested that these four branches were related to each other by more than geography, and referred to them all as Shoshonean (Lamb, 1964). Miller (1983, 1984), based on a review of cognate sets and comparisons of sound systems, rejected this relationship as genetic, preferring to view the four as independent branches of the family. Others accept the unity of these four branches, citing shared innovations in the sound systems and aspects of morphology as evidence (Goddard, 1996; Campbell, 1997; Heath, 1977, 1985; Manaster Ramer, 1992).

Internal diversity within the remaining branches of the family is also debated, with some arguing for and others against various subgroupings. Again, the problem of language extinctions and thus lack of data enters into the discussion (see Campbell, 1997: 133–135 for details), with Miller (1983) remarking that relationships make the family resemble less a tree than a vine that has been severely pruned! Names for the remaining branches also differ, but there is general agreement on Tepiman (Pimic), Taracahitan (Taracahitic), Corachol (Cora-Huichol), and Aztecan (Goddard, 1996; Campbell, 1997). The position of a fifth branch, Tubar, is likewise debated, with some placing it within Taracahitan (Kaufman, 1974). Some keep Tarahumara and Cahitan as independent branches (Ethnologue, 2004), and others include these with the first three named as a genetic subunit called Sonoran. Sonoran has a long history going back to Buschmann, with the most recent evidence being presented by Hale (1964). The unity of Southern Uto-Aztecan, including Sonoran and Aztecan, is less controversial than the proposal for Northern Uto-Aztecan (Campbell, 1997; Goddard, 1996; Heath, 1977, 1985; Manaster Ramer, 1992; Miller, 1983, 1984), although there is still some argument over the position of Aztecan as either having independent status within that unit or being more closely related to Corachol (Campbell and Langacker, 1978).

Most of the languages of Uto-Aztecan, with the exception of those that went extinct early, have been reasonably well studied, beginning with the work on Southern Paiute grammar, texts, and a dictionary by Sapir in 1910 (Sapir, 1930–1931). Most recent in a

long line of descriptive works is the publication of the massive *Hopi dictionary* (Hill *et al.*, 1998), representing the largest compilation to date for a Native American language. In between, sufficient descriptive works have been published by numerous authors to provide the data for reconstruction of the basic sound system of the proto-language, as well as an outline of some of its grammatical features, and a partial lexicon.

The Proto-Uto-Aztec sound system is considered by most to contain a single series of voiceless stops (p, t, c, k, k^w, ?), -s, h, two nasals [m, n (or ŋ)], a lateral (l), plus w, y, and possibly r, along with a five vowel system (i, a, i, o, u) plus vowel length (following Campbell, 1997). There is some disagreement as to the identity and directionality of the n, ŋ, and l: **n > *ŋ and **l > *n, particularly in selected environments in Northern Uto-Aztec, or **ŋ > *n and **n > *l in selected environments in Southern Uto-Aztec (see Campbell, 1997: 136–137) for discussion). The status of **r is likewise not clear, with some suggesting that it is one reflex of **t (Campbell, 1997: 137). Additional work with the cognate sets initially compiled by Miller (1967, 1988) may clarify the matter. The basic sound system has come down to the daughter languages with various alternations, with not all paths particularly clear.

Work on comparative grammar dates to the 1960s and 1970s (Heath, 1978; Langacker, 1977; Steele, 1979; Voegelin, Voegelin and Hale, 1962). Based on these studies, the proto-language is considered to have had several features, including an ‘absolutive’ noun suffix, used to mark a noun that is neither possessed nor carries another postposition; an auxiliary that contained a complex of modal, pronominal, and tense elements; and various pronominal elements on the verb that marked a reflexive object (Steele, 1979: 444–448). The proto-language is also considered to be a verb final language, with a much richer verb morphology than noun morphology.

The broad distribution of Uto-Aztec languages has spurred several investigations into the linguistic prehistory of the family, with archaeologists, anthropologists, and linguists making contributions through the years. Comparative lexical work has suggested homelands for Proto-Uto-Aztec in various locations within its present range, and various features, including agriculture, for its earliest speakers (Fowler, 1983; Hill, 2001, 2003). The language family has thus been fertile ground for testing many hypotheses from historical and theoretical linguistics to anthropological concerns. Today, the expertise of many Uto-Aztec specialists, especially in the United States, is also being given to Native communities in

partnerships addressing language salvage and revitalization in order to preserve these significant languages for speakers in the future.

Bibliography

- Campbell L (1997). *Oxford studies in anthropological linguistics 4: American Indian languages: the historical linguistics of Native America*. New York and Oxford: Oxford University Press.
- Campbell L & Langacker R (1978). ‘Proto-Aztec vowels: Parts 1, 2, and 3.’ *International Journal of American Linguistics* 44, 85–102; 197–210; 262–279.
- Ethnologue (2004). ‘Language family trees: Uto-Aztec.’ <http://www.ethnologue.com>.
- Fowler C S (1983). ‘Some lexical clues to Uto-Aztec prehistory.’ *International Journal of American Linguistics* 49, 224–257.
- Goddard I (1996). ‘The classification of the Native languages of North America.’ In Goddard I (ed.) *Handbook of North American Indians 17: Languages*. Washington, DC: Smithsonian Institution. 290–323.
- Hale K L (1964). ‘The sub-grouping of Uto-Aztec languages: lexical evidence for Sonoran.’ In XXXV *congresso International de Americanistas, México*, 2, 511–517.
- Heath J (1977). ‘Uto-Aztec morphophonemics.’ *International Journal of American Linguistics* 43, 27–36.
- Heath J (1978). ‘Uto-Aztec *na-class verbs.’ *International Journal of American Linguistics* 44, 211–222.
- Heath J (1985). ‘Proto-Northern Uto-Aztec principles.’ *International Journal of American Linguistics* 51, 441–443.
- Hill J H (2001). ‘Proto-Uto-Aztec: a community of cultivators in central Mexico?’ *American Anthropologist* 103, 913–934.
- Hill J H (2003). ‘Proto-Uto-Aztec and the northern devolution.’ In Renfrew C & Bellwood P (eds.) *Examining the farming/language dispersal hypothesis*. Cambridge, MA: McDonald Institute for Archaeological Research. 331–440.
- Hill K C, Sakaquaptewa E, Black M E, Malotki & the Hopi Tribe (1998). *Hopi dictionary*. Tucson: University of Arizona Press.
- Kaufman T S (1974). ‘Meso-American Indian languages.’ In *Encyclopedia Britannica*. 15th edn. Chicago. 11, 956–963. [Reprinted 1985 in *Encyclopedia Britannica* (15th edn.). Chicago. 22, 767–774.]
- Lamb S M (1964). ‘The classification of the Uto-Aztec languages: a historical survey.’ In Bright W (ed.) *Studies in California linguistics*. University of California publications in linguistics 34. Berkeley: University of California. 106–125.
- Langacker R W (1977). ‘An overview of Uto-Aztec grammar.’ *Summer Institute of Linguistics publications 56: studies of Uto-Aztec grammar I*. Arlington: SIL and University of Texas at Arlington.

- Manaster Ramer A (1992). 'A Northern Uto-Aztecan sound law: *-c-→ *y-.' *International Journal of American Linguistics* 58, 251–268.
- Miller W R (1967). 'Uto-Aztecan cognate sets.' *University of California publications in linguistics* 48. Berkeley: University of California Press.
- Miller W R (1983). 'Uto-Aztecan languages.' In Ortiz A (ed.) *Handbook of North American Indians 10: Southwest*. Washington, DC: Smithsonian Institution. 113–124.
- Miller W R (1984). 'The classification of Uto-Aztecan languages based on lexical evidence.' *International Journal of American Linguistics* 50, 1–24.
- Miller W R (1988). *Computerized data base for Uto-Aztecan cognate sets*. Salt Lake City: Department of Anthropology, University of Utah. [CD-R version by K C Hill University of Arizona, 2004.]
- Sapir E (1913–1914). 'Southern Paiute and Nahuatl: a study in Uto-Aztecan' (2 parts). *Journal de la Société des Américanistes de Paris*, n. s. 10, 379–425; 11, 443–488.
- Sapir E (1930–1931). 'The Southern Paiute language' (3 parts). Part 1: 'Southern Paiute, a Shoshonean language.' Part 2: 'Texts of the Kaibab Paiutes and Uintah Utes.' Part 3: 'Southern Paiute dictionary.' *Proceedings of the American Academy of Arts and Sciences* 65, 1–296, 297–535, 737–730.
- Steele S (1979). 'Uto-Aztecan: an assessment for historical and comparative linguistics.' In Campbell L & Mithun M (eds.) *The languages of Native America: historical and comparative assessment*. Austin: University of Texas Press. 444–544.
- Voeglin C F, Voegelin F M & Hale K L (1962). *Indiana University publications in anthropology and linguistics memoir 17: Typological and comparative grammar of Uto-Aztecan I: Phonology*. Bloomington: Indiana University Press.

Uyghur

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Uyghur (uyğur tili, uyğurça), formerly called Eastern Turki, is spoken in the Chinese Xinjiang Autonomous Region (Eastern Turkistan). It belongs to the South-eastern (Uyghur, Uyghur-Karluq, or Chaghatay) branch of the Turkic language family.

At least 10 million native speakers of Uyghur live in Xinjiang. This region borders Kirghizstan and Tajikistan in the west; Kazakhstan in the northwest; Mongolia in the north and northeast; the Russian Federation in the north; Afghanistan, Jammu, and Kashmir in the southwest; Tibet in the southeast; and the Chinese provinces Gansu and Qinghai in the east. Other languages spoken in Xinjiang include Chinese, Kazakh, and Kirghiz. The speakers of Uyghur predominantly live in the oases north of the Tarim river, on the southern slopes of Tianshan, and on the northern slopes of Kunlun in the southern Taklamakan desert up to the region west of the Lop desert. About half a million speakers of Uyghur live in eastern Kazakhstan, Kyrgyzstan, Uzbekistan, Tajikistan, Turkmenistan, Afghanistan, and Mongolia.

The status of Uyghur in Xinjiang is stable. Official documents are issued in both Uyghur and Chinese. Though education in Uyghur is possible up to the

university level, Chinese, the dominant language, is necessary for higher education. The Uyghurs make strong efforts to maintain and cultivate their language. In Kazakhstan, a few Uyghur schools exist, and there is a certain publishing activity in Uyghur.

Origin and History

The Old Uyghurs, living near the Selenga River on the territory of today's Mongolia, were vassals of the eastern Türk confederation. They defeated the Türk in 744 and created an empire that extended from Lake Baikal to the Altay Mountains. From here, they expanded their realm to Gansu in the east, and incorporated the Tarim basin and the Ferghana valley. The Uyghurs entertained close contacts with China, adopted Manicheism, and acted as the religion's protective power.

In 840, the Uyghurs were defeated by the Kirghiz, another Turkic-speaking tribal confederation. Most Uyghurs and many of their subjects fled southward. One group settled in the Ordos region of northern China and assimilated with Chinese and Mongols. The Uyghurs who settled in the Gansu corridor of western China, establishing contact with Tibetans and Mongols, are the ancestors of today's Yellow Uyghurs. The largest group fled to the Tarim basin. In Turfan, the southwesternmost possession of their steppe empire, the Uyghurs established the kingdom of Kocho, which expanded rapidly over large parts of the Tarim basin. It existed until the Mongol invasion in the 13th century, and as a semiautonomous state

for some time afterwards. A rich sedentary culture emerged, in which the Uyghur language was used for a comprehensive literary production.

Various Turkic groups had settled in this region from the 6th century on, particularly in the colonies of the Türk empire. The western Tarim basin was predominantly populated by Karluks. Large parts of the area had thus been Turkicized long before the Uyghurs settled here. The Turkic-speaking groups eventually absorbed the indigenous non-Turkic population of Indo-European origin, i.e., speakers of Soghdian (Sogdian) and Tokharian. The southern oases north of Kunlun were mostly populated by Saka speakers. Their region, which had its center in Khotan, essentially remained untouched by Turkicization up to the Mongol period.

The Old Uyghur culture was finally extinguished through the advancement of Islam. The first Islamic Turkic state in the east, the Karakhanid empire, was established in the 10th century, with Kashgar developing into the leading Islamic center in the east. Later on, the oasis states of the Tarim basin were controlled by Karakitay, Mongols, Junggars, and various local rulers. In the 20th century, Eastern Turkistan became a bone of contention in conflicts between Russia, China, and Britain.

Related Languages and Language Contacts

Uyghur is closely related to Uzbek. Modern Uyghur partly goes back to Old Uyghur, which is close to the language of the Orkhon inscriptions of the Türk dynasty. It is not a direct continuation of Old Uyghur, but differs considerably from it as a result of interaction with Indo-European varieties such as Soghdian and Tokharian, and other Turkic varieties. The Turkic varieties of Eastern Turkistan, a major crossroad of Central Asia, have been in contact with numerous other languages. A strong substratum influence has been exerted by speakers of Indo-European shifting to Turkic. Uyghurs had early contacts with Mongols, and later contacts with Kirghiz and Kazakhs. Elements of Persian and Arabic origin were spread by merchants and religious teachers along the Silk Road. Contacts with Russian began in the early 20th century. The long contacts with Chinese have been particularly important. The presence of Chinese speakers in Xinjiang has increased considerably since the 1950s.

The Written Language

Old Uyghur was a highly developed literary language, attested by rich historical materials, mostly of

a religious – predominantly Buddhist, but also Manichaeian and Nestorian Christian – nature. A rich treasure of Old Uyghur documents, written in various scripts, has been preserved, the last ones dating back to the 15th century. In the Islamic era, Eastern Turkistan was the birthplace of the literary language known as Karakhanid ('Khakani' Turkic), created in the 11th century in Kashgar, the cultural center of the Karakhanid state. Kashgar also became the basis of the eastern variety of the transregional literary language Chaghatay, which developed from the 15th century on in the Timurid realm. Modern written Uyghur is the last stage in this literary tradition.

In 1922, an assembly in Tashkent decided to adopt the historical term 'Uyghur' for speakers of Eastern Turki in Russian Turkistan. The designation was officially accepted in Xinjiang in 1934. Standard Uyghur, the official local language of Xinjiang, was originally based on varieties spoken in the Ili region, mainly on Soviet territory. The basis of the current standard language are the dialects of Ghulja and Ürümqi. Since 1954, the Language and Script Work Committee (Til-yëziq xizmiti komiteti) is responsible for its norms. The current standard pronunciation was determined in 1987 and slightly revised in 1997.

Modern Uyghur was first written with Arabic script. A Cyrillic script was introduced in 1957 but soon replaced by the Roman-based 'new script' (*yëngi yëziq*), based on the Chinese pinyin system. Since this experiment was unsuccessful, the Arabic-based 'old script' (*kona yëziq*) was revived in 1983. The return to it has made written communication with other Turkic-speaking groups more difficult. For Uyghur varieties of the Soviet Union, Arabic script was used up to 1930, then a Roman-based alphabet was used, and, from 1947 on, a Cyrillic script, which is still used in Kazakhstan, Kyrgyzstan, and Uzbekistan.

Distinctive Features

Uyghur exhibits most linguistic features typical of the Turkic family (see **Turkic Languages**). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation, e.g., *A* = *a/e*, *G* = *γ/g*, *K* = *q/k*. Segment within round brackets occur after consonant-final stems only. Hyphens are used here to indicate morpheme boundaries.

Phonology

Uyghur displays many features that are lacking or uncommon in other Turkic languages. The back

vowel *i* is missing in many environments where it is normally found in Turkic, e.g., *yil* ‘year.’ It is present in the neighborhood of back velars, e.g., *qiz* ‘girl’ (written *qiz*). The Arabic-based script does not distinguish the vowels *i* and *ï*, though it otherwise provides diacritic signs to designate vowels in an unambiguous way. The typical Turkic frontness-backness and rounded-unrounded harmony is preserved. The latter is relatively weak, many suffixes lacking variants with rounded vowels.

Certain vowel changes and phonological irregularities, in particular regressive assimilations, can be interpreted as Indo-European substratum phenomena. Unaccented nonhigh vowels are raised in open syllables, i.e., *a*, *e*, *é* > *i*, *o* > *u*, *ö* > *ü*, e.g., *bali-lar* ‘children’ (*bala* ‘child’), *yürüg-üm* [heart-POSS.1.SG] ‘my heart’ (*yürek* ‘heart’). Words of Arabic and Persian origin are not subject to this rule. Two rules of regressive assimilation may be due to contact with Iranian. First, *a* and *e* in open first syllables are raised to *é* under the influence of *i* or *ï* of the following syllable, e.g., *ét-im* [horse-POSS.1.SG] ‘my horse’ (*at* ‘horse’) or ‘my meat’ (*et* ‘meat’). Second, *a* and *e* in open first syllables are rounded under the influence of *u* or *ü* in the following syllable, e.g., *tömür* ‘iron’ < *temür*.

Finally -G is preserved in monosyllabic words, e.g., *tay* ‘mountain,’ but it is mostly changed to -K in nonfirst syllables, e.g., *tay-liq* [mountain-DER] ‘mountainous.’ Initial *ž*- occurring before high vowels corresponds to *y*- in many other Turkic languages, e.g., *yil* ‘year’ (Turkish *yıl*). The consonants *r* and *l* are often deleted, particularly before obstruents, e.g., *qa[r]* ‘snow,’ *qa[r]ga* ‘crow,’ *bo[l]sa* [be(come)-COND.3.SG] ‘if it is.’ Loanwords are restructured according to native phonotactical rules. Thus, *f* is mostly replaced by *p*, e.g., *pikir* ‘idea’ (<*fikr*). Non-permissible consonant clusters are broken up by means of epenthetic/prothetic vowels or consonant deletion, e.g., *xaliq* ‘people’ (<*xalq*), *iradiyo* ‘radio,’ *dos* ‘friend’ (< *dost*).

Grammar

The ablative suffix is -*Din*, as in Old Uyghur, whereas most Turkic languages exhibit -*Dan*, e.g., *öy-din* [house-ABL] ‘from the house.’ Uyghur has lost, as Chaghatay already had, the ‘pronominal *n*,’ which occurs, in most Turkic languages, in third person possessive suffixes before case suffixes, e.g., *öy-i-ge* [house-POSS.3.SG-DAT] ‘to her/his house’ (cf. Turkish *ev-in-e* [house-POSS.3.SG-DAT]). The polite form of the second person plural pronoun is *si-ler* [you-PL]. There is a general present tense marker going back to *-*a tur-ur* (converb suffix + ‘stands’), e.g., *oqu-y-du* [read-CONV.3.SG] ‘reads,

will read,’ and a more focal present marker going back to *-*p yat-a tur-ur* (converb + ‘lie’ + converb + ‘stands’), e.g., *oqu-wati-du* [read-FOCAL.PRES-3.SG] ‘is reading.’ Evidentiality is expressed with the copulas *éken/ëmiş* and the past suffix -(*i*)*p-tu*. The marker -*gan-di* expresses presumption. More than 20 auxiliary verbs are used in postverb constructions to express manner of action.

Lexicon

Uyghur displays numerous words of Arabic and Persian origin. Many words for abstract concepts are inherited from the Karakhanid-Chaghatay literary tradition. Though this influence has now decreased, one-fifth of the vocabulary is still of Arabic-Persian origin. A large part of the modern technical and administrative vocabulary has been copied from Russian. The lexical influence of Chinese has become increasingly stronger, and many Chinese neologisms have been copied. In the 1960s, the use of Chinese scientific terminology was obligatory. There is now a tendency to replace Chinese words with products of Turkic word formation, loan translations, and internationalisms copied from Russian.

Dialects

The classification of Uyghur dialects is still controversial. A northern group includes dialects spoken north and east of the Tianshan mountains and immediately south of them. It comprises the westernmost dialects of Kashgar and Yarkent, the more central dialects of Aqsu and Kucha, and the eastern dialects of Turfan and Qumul. The Kashgar-Yarkent dialect is strongly influenced by varieties of Western Turkistan. The Turfan dialect is of special interest, since it seems to stand in a direct historic relationship with Old Uyghur. A particular variety is Taranchi, spoken in the Ili valley by groups that emigrated from Eastern Turkistan at the end of the 18th century. This dialect, which is close to Uzbek, served as the basis of written Uyghur in Russian Turkistan. It is still spoken by Uyghurs in Kazakhstan, Kyrgyzstan, Uzbekistan, etc. A southern group comprises the Khotan dialects. A third group consists of the now nearly extinct Lopnur dialect, which was spoken the eastern Tarim basin and displayed Kirghiz and Mongolian influences. The Khoton (or Busurman ‘Muslim’) dialect is spoken in the region between the lakes Ubsu-nur and Chirgis-Nur in western Mongolia. The Eynu variety, spoken at various places in southwestern Xinjiang, combines an Uyghur morphosyntax with a special vocabulary of non-Turkic – partly Iranian and partly unknown – origin. Its speakers, all adult men,

use it as a secret language to make their conversations unintelligible to outsiders. Salar and Yellow Uyghur were formerly considered dialects of Uyghur. Salar is of Oghuz Turkic origin, whereas Yellow Uyghur is the continuation of an Old Uyghur dialect.

Bibliography

- Dwyer A M (2001). 'Uyghur.' In Garry J & Rubino C (eds.) *Facts about the world's major languages: an encyclopedia of the world's major languages, past and present*. Dublin: The H. W. Wilson Company/New York: New England Publishing Associates. 786–790.
- Hahn R F (1991). *Spoken Uyghur*. Seattle: University of Washington Press.
- Hahn R F (1998). 'Uyghur.' In Johanson L & Csátó É A (eds.) *The Turkic languages*. London: Routledge. 379–396.

- Jarring G (1933). *Studien zu einer osttürkischen Lautlehre*. Lund: Borelius, Leipzig: Otto Harrassowitz.
- Jarring G (1946–1951). *Materials to the knowledge of Eastern Turki* 1–4. Lund: Lunds Universitets årsskrift.
- Jarring G (1964). *An Eastern Turki-English dialect dictionary*. Lund: C. W. K. Gleerup.
- Menges K H (1955). *Glossar zu den volkskundlichen Texten aus Ost-Türkistan*. Wiesbaden: Harrassowitz.
- Nadzhip E N (1971). *Modern Uigur*. Moscow: Nauka.
- Pritsak O (1959). 'Das Neuigurische.' In Deny J et al. (eds.) *Philologiae turcicae fundamenta* 1. Aquis Mattiacis: Steiner. 525–563.
- Schwarz H G (1992). *An Uyghur-English dictionary*. Bellingham: Western Washington University Press.
- Wei C Y (1989). 'An introduction to the Modern Uygur literary language and its dialects.' *Wiener Zeitschrift für die Kunde des Morgenlandes* 79, 235–249.

Uzbek

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Uzbek (*ozbek tili*, *ozbekcha*) belongs, like modern Uyghur, to the southeastern (Uyghur, Uyghur-Karluk, or Chaghatay) branch of the Turkic language family. It is spoken in various dialects in Western Turkistan, primarily in the Republic of Uzbekistan (*Ozbekiston Respublikasi*), which occupies the major part of Transoxiana and has common borders with Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, and Turkmenistan. The Uzbek-speaking areas are concentrated in the the lower Zerafshan and upper Syrdarya valleys and in the Ferghana valley, west and north-west of western Tienshan. Though the Uzbeks make up 80% of the population of the republic, Uzbek is spoken by less than 75%, i.e., about 19.8 million people. Other languages of Uzbekistan include Russian and Tajik. The latter is mainly spoken in the oases of Bukhara and Samarkand and in the Ferghana valley. Russian is mainly spoken in the capital, Tashkent. Karakalpakstan, which comprises the northwestern part of Uzbekistan, is an autonomous republic with a special status and a language of the Kazakh type. Karakalpaks make up 2% of the population. Many Karakalpaks use Uzbek as a second language. Uzbek is also spoken in parts of Tajikistan (1.2 million speakers), northern Afghanistan (1.5 million), Kyrgyzstan (600 000),

Kazakhstan (350 000), Turkmenistan (360 000), and China (Xinjiang) (15 000). The total number of speakers amounts to about 24 million people.

Uzbek–Russian bilingualism is widespread in Uzbekistan, and Russian has a strong position. Uzbek is, however, one of the most firmly established Turkic languages, the second after Turkish in terms of cultural importance. It has been subject to systematic language planning and cultivation. Post-Soviet Uzbek is in a transitional period of dynamic developments. In general, the status of the Russian language has declined. In the first post-Soviet years, Russian was still defined as the medium of 'crossnational communication' in Uzbekistan. Later, it lost this role and its status as a compulsory subject in Uzbek education. However, the goal of enforcing obligatory use of the indigenous languages in public functions within a few years' time has proved unrealistic.

Origin and History

The historical background of the current linguistic situation is highly complex. In what is now Uzbekistan, varieties of southeastern Turkic have been spoken for a millennium, both by nomadic groups and by a sedentary population in close contact with Iranian-speaking groups. An intensive Iranian–Turkic bilingualism developed in Transoxiana and Ferghana. Sizeable Iranian-speaking groups eventually shifted to Turkic. The Uzbek conquest brought in a different element (the original Uzbeks spoke a Kipchak language of the Kazakh type). After the fall

of the Golden Horde, the Uzbeks left the Kipchak territory in the Ponto-Caspian area, seized the power in Transoxiana in about 1500, overthrew the Chaghatay empire, and established the khanates of Khiva and Bukhara. The language of these politically dominant groups was gradually absorbed by varieties of southeastern Turkic ('de-Kipchakization'). Different degrees of maintenance of Kipchak elements and of Iranicization mirror the transition from nomadic to sedentary life. Modern Uzbek is based on the language of the old sedentary Turkic population but displays a few Kipchak features. The original Kipchak varieties have gradually vanished with the abandonment of the nomadic lifestyle. Some remnants of them are found in northern and northwestern Uzbekistan. Russia conquered Uzbekistan in the late 19th century. In the 1920s, Soviet Turkistan was split into a number of socialist republics, and an Uzbek republic was set up in 1924.

Related Languages and Language Contacts

Uzbek is most closely related to modern Uyghur, with which it shares important features. The oldest Turkic population of the area had close relations to speakers of Iranian. The predecessor of Uzbek was strongly influenced by Sogdian and, after the Muslim conquest, New Persian. Long-standing intensive contacts with East Persian have led to copying of numerous features in phonology, morphology, vocabulary, and syntax. There are many striking structural similarities between Uzbek and Tajik. These influences, along with later Kipchak Turkic and Russian influences, have given Uzbek a highly composite character. Uzbek–Tajik bilingualism is still alive in some areas, e.g., in and around Samarkand and Bukhara. An increasing Uzbek influence on Tajik may be observed. Kirghiz–Uzbek bilingualism is found in Kirghiz towns bordering on the Uzbek Ferghana valley.

The Written Language

Persian was for a long time the prestigious language of administration and higher culture in Transoxiana. Its importance later decreased in favor of Chaghatay, the Persian-influenced literary language of the Chaghatay empire cultivated in Samarkand, Bukhara, Tashkent, Ferghana, and other centers. From the 18th century on, Chaghatay developed further through successive modernization and adaptation to regional spoken varieties. The 'Sart' literary language, used until 1920, consisted of Chaghatay with certain modern regional elements. After the

Russian conquest of the region, Uzbek was established as the standard language. It was first based on northern dialects, and later, after 1937, on the southern dialects of Tashkent and Ferghana. Modern standard Uzbek is the common 'roof' of highly different varieties.

Uzbek was first, as Chaghatay, written in Arabic script. A Roman-based script was introduced in 1929 and revised in 1934. In 1937, the orthography was simplified. Due to this reform, Uzbek texts became less easily intelligible to readers in neighboring Turkic areas. The old role of Uzbek as a transregional language was thus drastically restricted. A Cyrillic-based script was introduced in 1940 and 1941 and was later modified. The transition to a Roman-based alphabet was enacted by law in 1993. The new script system was revised in 1995. It is based on the American Standard Code for Information Interchange (ASCII) and dispenses with diacritic signs. The new system preserves the principles of the Cyrillic-based system. It essentially represents a transliteration of the Cyrillic spelling and is thus very different from the Roman-based system of the 1930s.

Distinctive Features

Uzbek exhibits most linguistic features typical of the Turkic family (*see Turkic Languages*). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following discussions, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation, e.g., A = a/æ, G = γ/g, K = q/k. Hyphens are used to indicate morpheme boundaries.

Phonology

The phonetic realizations of the vowels vary greatly. The standard orthography is very vague about the actual pronunciation. With the revision of the Roman-based script in 1934, the vowel signs were reduced to six. When, in 1937, the strongly Iranicized Tashkent dialect was chosen as the norm for the standard language, the signs for *ö*, *ü*, and *ï* disappeared, and the front vowel *æ* was written with the letter 'a.' These principles have been maintained in the new Roman-based system. Modern spelling thus applies a system of six vowel signs, identical with the system used for Tajik. It does not reflect the fact that the distinctions between back and front vowels have been largely preserved.

Standard Uzbek is claimed to have, as a result of Iranian influence, the six vowel phonemes *a*, *e*, *ä*, *i*, *o*, and *u*. This analysis is mirrored in the orthography. The sign 'a' stands for a front *æ*, but also for a backed

‘a’ when adjacent to back-velar or uvular consonants. The higher vowel *e* occurs in first syllables, e.g., *er* ‘husband’ and *keræk* ‘necessary.’ A labialized *â* occurs in first syllables, e.g., *âlti* ‘six’ and *âra* ‘interval,’ confusingly enough written with the letter ‘o.’ Similarly, the letter ‘i’ stands for a front *i*, but also for a backed *ï* when adjacent to back-velar or uvular consonants, e.g., *yaxši* ‘good.’ High, unrounded vowels are often reduced or lost in closed syllables before certain consonants, e.g., [b^ɪr] ‘one.’ The vowels *ô* and *û* (with a somewhat retracted pronunciation) occur instead of Common Turkic *ö* and *ü*, e.g., *ôlik* ‘dead’ and *tûn* ‘night.’ The distinctions *o* vs. *ô* and *u* vs. *û* are not reflected in the script. Thus, pairs such as *bol-* ‘become’ vs. *bôl-* ‘divide’ and *uç* ‘end’ vs. *ûč* ‘three’ are homographic in modern spelling.

Due to Iranian influence, the manifestations of sound harmony are less straightforward than in most other Turkic languages. There are disturbances of the vowel harmony in most urban dialects, whereas the northern dialects have preserved vowel harmony. Suffixes are very often invariable and their vowels are not assimilating to the frontness–backness or the roundedness–unroundedness of the preceding vowel.

The notation of consonants in the new orthography follows the principles of the Cyrillic-based script. For example, ‘ch’ represents *č*, ‘sh’ represents *š*, and ‘j’ represents *ǰ*. The back velars *q* and *ɣ* are represented by *q* and *g*; the front velars *k* and *g* are represented by ‘k’ and ‘g’. As in Uyghur, final -G has mostly changed to -K in nonfirst-syllable positions, e.g., *sariq* ‘yellow’ and *tirik* ‘alive’ (cf. Turkish *sarı* and *diri*).

The realizations of suffixes are highly regular. Uzbek displays less consonant assimilations of suffix-initial *n*, *d*, and *l* than most neighboring languages do. The standard spelling is basically morphological and normally does not indicate vowel harmony or consonant assimilations. In loanwords, high epenthetic vowels are inserted to dissolve nonpermissible consonant clusters, e.g., *fikⁱr* ‘thought’ (written *fikr*). However, elements of Russian, Arabic, and Persian origin usually reflect the original forms as written in Cyrillic and Arabic script. Assimilations and other adaptations are thus obscured, e.g., *nisbat* ‘relation’ for [*nispæt*].

Grammar

Uzbek lacks the ‘pronominal *n*’ in the declension of nouns with third-person possessive suffixes, e.g., *qol-i-da* [hand-POSS.3.SG-LOC] ‘in his/her hand’ (Turkish *kol-un-da* [hand-POSS.3.SG-LOC]). The genitive suffix -*ni* is mostly pronounced as -*ni*, thus coinciding with the accusative suffix. Like Uyghur, Uzbek has abandoned the old pronominal flexion in favor of nominal

flexion, e.g., *sen-gæ* [you-DAT] ‘to you’ (Turkish *sana* [you-DAT]). The demonstrative pronouns constitute a four-place system with *bu* ‘this,’ *šu* ‘this (in view),’ *oša* ‘this (in view, more distant),’ and *u* ‘that.’ Lower numerals can assume the suffix -*tæ*, e.g., *ikki-tæ* ‘two (pieces).’ The suffixes -*âw* and -*ælæ* form collective numerals, e.g., *ikk-âw* ‘the two’ and *ikk-ælæ-si* [two-COLL-POSS.3.SG] ‘both of them.’

The general present-tense marker goes back to converb suffix + ‘stands,’ e.g., *kel-æ-di* [come-PRES-3.SG] ‘comes, will come.’ More focal present markers include -(*æ*)*yæp* and -(*æ*)*yâtir*, going back to constructions with *yât-* ‘to lie,’ e.g., *yâz-yæp-ti* [write-FOCAL.PRES-3.SG] ‘is writing (just now)’ and *kel-æ-yâtir* [come-FOCAL.PRES-3.SG] ‘is coming.’ Other focal forms with various nuances can be formed with auxiliary verbs meaning ‘to stand,’ ‘to sit,’ or ‘to move,’ e.g., *yâz-ib turib-mæn* [write-FOCAL.PRES-1.SG] ‘I am writing.’ Evidentiality is expressed by the indirective past marker -(*i*)*bdi* and the indirective copula particles *ekæn/emiš*, e.g., *ayt-ib-di* [say-EV-3.SG] ‘obviously said,’ *unut-ib-mæn* [forget-EV-1.SG] ‘I have obviously forgotten,’ and *kæsæl ekæn* [ill COPEV] ‘is obviously ill.’ The interrogative forms -*mi-kæn* and -*mi-kin* express doubt, e.g., *kel-gæn-mikan?* [come-POSTTERMINAL.PAST INTERROG] ‘has (s)he really come?’ The postterminal (‘past’) marker -*gæn* is used as participle and as a finite form, as in most other Turkic languages, corresponding functionally to Turkish -*mlş*, -(*y*)*An*, and -*DIK*. There is also an intraterminal (‘present’) participle in -*digæn*. Postverb constructions with converb forms of the lexical verbs plus auxiliary verbs are used to express semantic modifications, including manner of action, e.g. -*æ ber-* ‘to do continuously, to keep doing.’ The Persian impact on Uzbek syntax is considerable.

Lexicon

The Uzbek vocabulary contains many loans of Arabic–Persian origin, mostly copied from Persian and inherited from the old literary language, Chaghatay. The strong Iranian influence on Uzbek has led to borrowing of word-formation affixes, even prefixes, e.g., *nâ-toʻyri* ‘untrue’ (*nâ-* ‘non-’ copied from Persian plus Turkic *toʻyri* ‘right, true’). Female gender is expressed in some nouns borrowed from Arabic and Russian, e.g., *šâir-a* [poet-FEM] ‘poetess’ (*šâir* ‘poet’) and *student-ka* [student-FEM] ‘female student’ (*student* ‘student’). Most conjunctions are of Arabic–Persian origin.

Dialects

The dialects exhibit different degrees of Iranicization. The northern dialects, spoken in southern Kazakhstan, north of Tashkent, show little Iranian

influence. Southern Uzbek includes the dominant urban dialects of Tashkent, Bukhara, Samarkand, etc., which go back to varieties of the settled populations. They have been heavily influenced by East Persian in their vowel system, for example. Moderately Iranicized dialects are spoken east of Tashkent, in the Ferghana valley, representing a successive transition to Uyghur. The rural dialects of the Kipchak type are Kazakh dialects. Oghuz Turkic dialects, improperly called Oghuz Uzbek, are spoken in Khorezm and adjacent areas. Related dialects are also spoken in southern Uzbekistan and in southern Kazakhstan.

Bibliography

- Boeschoten H (1998). 'Uzbek.' In Johanson L & Csátó É Á (eds.) *The Turkic languages*. London & New York: Routledge. 357–378.
- Borovkov A K (ed.) (1959). *Uzbeksko-russkij slovar'*. Moskva: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovarej.

- Gabain A von (1945). *Özbekische Grammatik*. Leipzig: Otto Harrassowitz.
- Johanson L (2001). 'Uzbek.' In Garry J & Rubino C (eds.) *Facts about the world's major languages: an encyclopedia of the world's major languages, past and present*. New York, Dublin: The H. W. Wilson Co., New England Publishing Assoc. 791–793.
- Kononov A N (1960). *Grammatika sovremennogo uzbekskogo jazyka*. Moskva, Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Raun A (1969). *Basic course in Uzbek*. Bloomington: Indiana University/The Hague: Mouton.
- Sjoberg A F (1963). *Uzbek structural grammar*. Bloomington: Indiana University/The Hague: Mouton.
- Waterson N (1980). *Uzbek-English dictionary*. Oxford: Oxford University Press.
- Wurm S (1959). 'Das Özbekische.' In Deny J *et al.* (eds.) *Philologiae turcicae fundamenta 1*. Aquis Mattiacis: Steiner. 489–524.

Relevant Website

- <http://www.turkiclanguages.com> – Website with many Turkish language resources.

V

Vietnamese

J Edmondson, University of Texas at Arlington,
Arlington, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

Historical Origins

The Vietnamese are thought to be descended from a precursor people who once dwelled near the Viet-Lao border of Central and Northcentral Vietnam where today are still found Vietic groups such as the Mu'òng, Nguồn, Arem, Rục, Pọng, Arem, Mày, Sách, Mã Liềng, and Thà Vù'ng (Nguyễn Tài Cẩn, 1995), (Ferlus, 1979, 1982, 1991, 1996, 1997). Long ago some of these groups moved north into the Red River Valley, lived under their own Hùng Vu'òng kings, and then allied with the indigenous Tày ethnicities in the joint Tày-Viet Kingdom of Âu-Lạc at Cổ Loa Citadel near Hanoi (257 to 208 B.C.). This lineage ended when the First Emperor of China and builder of the Great Wall dispatched the Chinese general 趙佗 Zhao Tuó (in *Hanyu Pinyin* transcription) or Triệu Đà (in Vietnamese), who conquered Âu-Lạc and introduced 100 000 soldiers, Chinese rule, Chinese character writing or 𡵚漢 *Chũ' Hán* 'Han (Chinese) script', and the Mandarin administrative system.

Vietnam remained a Chinese province until A.D. 939. After the Chinese departed, the Vietnamese continued the practice of borrowing from the Chinese cultural lexicon as well as structural and grammatical forms, and continued developing the character script, which was then renamed 𡵚儒 *Chũ' Nho* or 'learned script', and which contrasted with newly created characters used for writing purely Vietnamese lexical items. This new demotic script was called 𡵚喃 *Chũ' Nôm* 'southern, local, vernacular script' or simply *Nôm*, which was clearly in evidence by the 13th century, but was perhaps in use as early as the 10th century. Consider the following examples of strategies used for crafting *Nôm* characters: (1) 𡵚 *giò'i* 'heaven, sky' from Chinese 天 meaning 'heaven' and 上 meaning 'above'; (2) 𡵚 *đât* 'earth' composed of 土 *thổ* 'earth' for the meaning and 旦, taken from one half of 𡵚 for the sound *đât*; (3) 𡵚 *cá* 'fish' is a

combination of 魚 'fish' for the meaning and 𡵚 *cá* for the sound; and (4) 𡵚 *ba* 'three' is composed of the radical 三 meaning 'three' and 巴 *ba* for the sound. These examples show that 10th century Vietnamese were very much aware of the principles used in the construction of Chinese characters, namely to combine a radical part for meaning and a phonetic part for the sound, but the example for *sky*, *heavens* shows that sometimes other methods of creation were used.

Chinese influence in Vietnamese is generally very important and is the result of (1) 1000 years of occupation by Chinese speakers, (2) the role of Chinese as the spoken and written language of administration and, (3) the fact that Chinese continues to be the source of borrowing even today. Chinese loans in contemporary Vietnamese, called Sino-Vietnamese, can make up as much as 80% of the vocabulary in some semantic domains, (Hoàng, 1991: 5). But the depth of Chinese influence extends beyond the lexical. Indeed, some of the typological incongruities of morphology and syntax are now considered to be the result of contact with Chinese and other languages.

Politically independent at last, the Vietnamese then turned their attention to southern rivals, the Champa Kingdom. Three hundred years of greater and lesser hostilities ensued, with ebbs and flows evident in marriage alliances and other accommodations; Vietnamese absorbed some early loans from this source as well. In the 14th century, the Vietnamese gained ultimate control and the boundaries of the language were advanced southward until Vietnam reached its current geographic extension. Chinese *Chũ' Hán* with an admixture of *Chũ' Nôm* writing flourished until the late 1600s when an outside force, Jesuit missionaries including Alexander de Rhodes and Francisco de Pina, developed an orthography based loosely on Portuguese and on Italian models that were intended not for the court but for believers among the common people. It was a romanized script with diacritics for tones and vowels, called *Chũ' Quốc Ngữ*, whose timeline can be sketched as follows: 1620–1631 embryonic beginnings, 1631–1648 revisions, 1651–1659 dissemination, and 1772–1838 finalizing stage, (Lý, 1999: 234–5).

Over the 18th and 19th centuries, official character script and popular roman script co-existed, but gradually *Chữ Quốc Ngữ*, aided first by French colonial proclivities in favor of a Latin-based script and the Church and later by populist movements, led to a shift in orthography; the character-based writing of Vietnamese was finished entirely by 1917 when the French eliminated the Chinese examination system, cf., Alexandre de Rhodes (1651) and Nguyễn Đình Hoà (1973, 1992).

Regional Varieties

The regional varieties of Vietnamese are divided into three main types: Northern (e.g., Hanoi), Central (e.g., Huế), and Southern (e.g., Hồ Chí Minh City). Recently, it has been suggested by Ferlus and Nguyễn Tài Cẩn that a fourth regional area should be added, North-central Vietnam or Area IV (Nghệ An Province), as this area preserves several special features lost elsewhere, cf., below and Alves (2000).

Phonological Forms

Vietnamese of the northern type has six tones, the southern type five, and central/northcentral types five or six, all of which can be traced back to a parent tone system of three columns of *level*, *rising*, and *falling* tones and two rows, *high* and *low*. Haudricourt (1954) proposed that Vietnamese was originally not a tonal language but that tonality arose as a part of the historical development of Vietnamese. In this theory syllable-final consonants caused changes of pitch; rising tones were produced in syllables that formerly ended in -p, -t, -k, -ʔ, whereas falling tones were created in syllables that once ended in -s or -h, and mid-level tones were created when syllables possessed no final consonants to draw the pitch up or down. Haudricourt's famous theory of Vietnamese *tonogenesis* explains how a non-tonal Mon-Khmer language could change its typological features and become tonal. Vietnamese tone categories are also associated with specific voice quality contrasts that accompany each tone, perhaps a residue of its tonogenetic history. Thus the tone called *ngang* demonstrates mid-level with modal voice (tones are notated with the scale-of-five system in which 5 is the highest and 1 is the lowest level, cf. Y.-R. Chao, 1930); *huyền* falls from mid with lax/breathy voice; *sắc* rises sharply from mid with tense voice ending in a glottalized coda; *nặng* falls with increasingly tense voice from early in the syllable to glottal stasis; *hỏi* is a fall-rise tone; *ngã* has a glottal interruption in the center of the syllable, sounding almost as if it were composed of two syllables VʔV, with overall a very high rising pitch, cf., Nguyễn and Edmondson (1997). The

names *ngang*, *huyền*, etc., illustrate the tone they name. See Table 1.

Vietnamese consonants in Hanoi speech distinguish five places of articulation – *labio-dental*, *denti-alveolar* (the *t* series is denti-alveolar initially and apico-postalveolar finally), *palatal*, *velar*, and *glottal*, and several manners of articulation – *voiceless aspirated stops*, *voiceless unaspirated stops*, *preglottalized voiced stops*, *fricatives* (*x* is a lamino-prepalatal narrow grooved fricative), *liquids*, and *nasals*. Compare *Chữ Quốc Ngữ* and IPA values of these consonants in Table 2.

In southern Vietnamese there are some important differences in initials compared to Hanoi. Notably *tr-* and *s-* are retroflexed [tr̠ʂ] and contrast with *ch-* [tɕ] (with very little friction compared to Hanoi speech) and *x* [s] respectively, *g-* [g], *r-* [r] (with a lot of variation in realization of the *r-*), *v-* [j] (Thompson, 1987: 89) *d-*, *gi-*[j]. As mentioned above, Northcentral Vietnamese has preserved three distinct pronunciations for *d-*, *r-*, and *gi-*. See Table 3.

Vietnamese word structure allows only sequences C₁V(V)C₂, whereby C₁ can be any allowed initial (notice that *p-* cannot be an initial except in words borrowed from French, e.g., *pín* ‘battery’ and *píp* ‘smoking pipe’). The syllable coda C₂ can be -*p̚* -*h̚* -*ʔ* -*k̚* -*ʔch* -*m* -*n* -*nh* [ŋ] -*ng* [ŋ]. One noteworthy phonological change in northern speech is engendered by these final consonants. Whenever the velars -*c* or -*ng* follow back rounded vowels *u*, *ô*, or *o*, there is *double closure*, and a velar and labial are simultaneously formed (because these are also accompanied by glottal stop, in reality they are triply closed). The preceding round vowel causes simultaneous assimilatory rounding of [-k̚ -ŋ] to [-k̚p̚ -ŋ̚m̚] in addition to diphthongization of the vowel, e.g., *dùng* ‘to use’

Table 1 Vietnamese tone categories

| | <i>Level</i> | <i>Rising</i> | <i>Falling</i> |
|-------------|------------------|----------------------------|--------------------------|
| <i>High</i> | ngang [ŋaŋ33] | sắc [săk35 ⁵] | hỏi [hỏi323] |
| <i>Low</i> | huyền [hwiən 31] | nặng [năŋ21 ¹] | ngã [ŋa4 ² 5] |

Table 2 Vietnamese consonant system

| | | | |
|---------|-----------------|------------------|-----------|
| ph- [f] | th- [tʰ] | kh- [x] | |
| -p [p̚] | t [t̚] | c, k, q- [k] | |
| | | tr-, ch [tɕ] | |
| b- [b] | đ [ʔd] | g(h)- [g] or [ɣ] | |
| | x-, s- [s] | | h- |
| v- [v] | r-, d-, gi- [z] | | |
| m [m] | n [n] | nh [ɲ] | ng(h) [ŋ] |
| | l- [l] | | |

[z^uŋm³¹] *duc* a suffix [z^uŋk²¹], *học* ‘study, -ology’ [ha^uk²¹] and *đông* ‘Vietnamese piastre’ [ʔd^uŋm³¹]. Similarly, whenever the palatals *-ch* or *-nh* follow the unrounded vowels *-i*, *-ê*, or *-a*, then diphthongization occurs as before, but there is no rounding, as all segments are unrounded to begin with, e.g., *minh* ‘clear, bright’ [mi¹ŋ33], *thích* ‘to like’ [thi¹c35²], *lệnh* ‘order, command’ [lɔ¹n²¹], *ếch* ‘frog’ [ɛ¹c35²], *anh* ‘Sir, you, older brother’ [ɔ¹ŋ33] and *thạch* ‘stone’ [thɜ¹c21¹].

The rhymes of Vietnamese syllables can have the nuclear vowels: /i iə e ɛ ɛː w u ʌ o ɔ ɔː/. If one assumes that /iə w ʌ u ʌ/ function as the long versions of /i u u/ u/, then all vowels except /e/ have long and short forms. In Table 4 one sees the possible

combinations of these nuclear vowels with the set of possible codas. Nuclear vowels may occur in open syllables, i.e., with *-ə* coda or in one of the following combinations: *-j* (graphically *-i/-y*), *w* (graphically *-o/-u*), *-m*, *-p*, *-n*, *-t*, *-nh*, *-ch*, *-ng*, *-c*. There are some notable areas where combinations are disallowed. For example, there are no rhymes *ij, *uw, or *ow, palatal coda may combine only with *-a*- and *-i*-, and velars may not combine with high front vowels except /ɛː/.

Word Category and Constructions

Contrary to some reports, Vietnamese is not an absolutely monosyllabic language, but one with many compounds and reduplicated structures. Compounds demonstrate disyllabic construction and have either been borrowed intact from Chinese (so nouns are right-headed, e.g., *cổ đại* [old-[era_N]_N] ‘antiquity’, *thanh âm* [voice-[sound_N]_N] ‘sound’, and verbs are left-headed *ẩu thổ* [v_V vomit]-spit] ‘to vomit’) or are pure Vietnamese creations of Vietnamese or mixed lexical roots, all left-headed *ngủ’o’i Việt* [N_N people]- Viet] ‘Vietnamese people’, *nhà thu’o’ng* [N_N house]-of the injured] ‘hospital’, *làm việc* [v_V do]-work] ‘to work’, except for the group of *father-mother compounds*, which consist mostly of semantically paired things, e.g., *bố mẹ* father-mother ‘parents’, *bàn ghế* table-chair ‘furniture’, *bát đĩa* bowl-plate ‘dishes’, which function as a unit without head. Vietnamese reduplicatives are also very productive, such as *complete reduplicatives* having the same onset and rhymes, *cười’i-cười’i* ‘laugh a little’, *nói nói* ‘keep talking’, *register change reduplicatives* with the repeated part, be it on the right or left, having opposing

Table 3 Regional variation of initials from Alves (2000). QN = Quốc Ngữ, NV = Hanoi, NCV = locations in Nghệ An Province, CV = Huế, SV = Hồ Chí Minh City

| QN | NV | NCV | CV | SV |
|-----|----|-----|-----|-----|
| s | s | ʃ | ʃ | ʃ |
| x | s | s | s | s |
| tr | c | ʈr | ʈr | ʈr |
| ch | c | c | c | c |
| r | z | r | r | r |
| d | z | j | j | j |
| gi | z | z | j | j |
| v | v | v | j | j |
| -nh | ŋ | ŋ | n | n |
| -n | n | n | n/ŋ | n/ŋ |
| -ng | ɲ | ɲ | ŋ | ŋ |
| -ch | c | c | t | t |
| -t | t | t | t/k | t/k |
| -c | k | k | k | k |

Table 4 Vietnamese rhymes with nuclear vowel(s) in the left column and codas in rows. Compiled from Lê Văn Lý (1948), Haudricourt (1951), Gordina (1960), Emeneau (1951), as reported in Cao Xuân Hạo (2003: 88–103)

| | Ø | j -i/y | w -o/u | m -m | p -p | n -n | t -t | ɲ -nh | c -ch | r -ng | k -c |
|----|-----|-----------|-----------|---------|---------|---------|---------|----------|----------|----------|---------|
| i | i | - | iu | im | ip | in | it | inh | ich | - | - |
| iə | ia | - | iêu | iêm | iêp | iên | iêt | - | - | - | - |
| e | ê | - | êu | êm | êp | ên | êt | - | - | - | - |
| ɛ | e | - | eo | em | ep | en | et | - | - | - | - |
| ɛː | e | - | - | - | - | - | - | - | - | eng | ec |
| ʌ | u’ | u’i | u’u | - | - | - | u’t | - | - | u’ng | u’c |
| ʌ | u’a | u’o’i | u’ou | u’om | u’op | u’on | u’ot | - | - | u’ong | u’oc |
| ʏ | - | ây | âu | âm | âp | ân | ât | - | - | âng | âc |
| ʏː | o’ | o’i | - | o’m | o’p | o’n | o’t | - | - | - | - |
| a | - | ay | au | ăm | ăp | ăn | ăt | anh | ach | ăng | ăc |
| aː | a | ai | ao | am | ap | an | at | - | - | ang | ac |
| u | u | ui | - | um | up | un | ut | - | - | ung | uc |
| uʌ | ua | uôi | - | uôm | - | uôn | uôt | - | - | uông | uô |
| o | ô | ôi | - | ôm | ôp | ôn | ôt | - | - | ông | ôc |
| ɔ | o | oi | - | om | op | on | ot | - | - | ong | oc |
| oː | - | - | - | - | - | - | - | - | - | ông | - |
| ɔː | - | - | - | - | - | - | - | - | - | oong | ooc |

high-low register of the same tone class, e.g., *xẹp* 'be flattened' vs. *xếp-xẹp* 'be completely flattened' (with a sắc tone syllable being here followed by a nặng reduplicant); some have vowel changes in the reduplicant, e.g., *hốc* 'hole, hollow' vs. *hốc-hác* 'be emaciated, gaunt'; some can have changes of onset in the first element and some in the second element as *rộn* 'be noisy' vs. *chộn-rộn* 'be agitated' and *xẹp* 'be flattened' vs. *xẹp-lép* 'be completely flattened', (Thompson, 1987).

Regarding word categories, Vietnamese has the following: nouns *Lê Quý Đôn* 18th century author, *học sinh* 'student', *gà* 'chicken'; classifiers *hai con chó* two-CLS-dog 'two dogs', *các cái bàn* plural-CLS-table 'tables'; locatives *ngoài* 'outside', *bắc* 'north'; numerals *hai* 'two', *ba* 'three'; verbs *đi* 'to go, walk', *nói* 'to talk'; stative verbs (adjectives) *tốt* 'good', *mới* 'new'; and pronouns. It is noteworthy for its rich and complex set of pronouns. There are personal pronouns *tôi* 'I, me (with modesty, servant)', *ta* 'I (emphatic), one', *tao* 'I (arrogant)', *mày, mi, bay* 'you (arrogant)', and *nó* 'it (animal)' or 'he (for children or contemptible persons, criminals)'. The term *mình*, meaning 'body' is used for 'you (intimate)'. The term *chúng* 'group of animate objects' can be combined with the above to make plurals such as *chúng tôi* 'we-exclusive' and *chúng ta* 'we-inclusive'. In public discourse, kinship names are often used, e.g., *chị* 'older sister, you-Ms.', *bà* 'grandmother, old woman, you-Madame', *anh* 'older brother you-Sir', *cháu* 'niece/nephew, grandchild, you-Young Person', *cô* 'father's sister, you-Ms.' (One speaker from Hanoi said that *cô* was obligatory to address one's female teachers.) These can become like 3rd person pronouns by adding *ấy*, e.g., *anh ấy* 'he, that Sir (a contemporary)', *chị ấy* 'she, that Ms. (a contemporary)', but there is also *nó* 'he (deprecating)' and *họ* 'they'. Kinship names also have features of anaphoric nouns, they contain additional information about gender and degree of familiarity, and they function differently in tracking participants in discourse.

Phrases and Sentences

Phrases are mostly left-headed, e.g., attributive adjectives follow heads, *đồng Việt Nam* piaster-Vietnam 'the Vietnamese piaster', complements follow heads, *ăn cơm* eat-rice 'to eat (food)', whereas adverb-like elements can appear to the left or right of the head, *rất đắt* very-expensive 'very expensive' but *đắt lắm* expensive-very 'very expensive'.

Sentences tend to have known, presupposed information at the beginning and new, asserted information at the end. One manifestation of this principle is that

after introducing a referent, a close-knit group of clauses or a topic chain follows whose subjects are PRO, the zero pronoun. For example in the famous story about Bần with no overt pronouns, one finds:

Bần chỉ là một anh nghèo xác, e_i ngày ngày lang thang khắp xóm này khác e_i xin ăn. Bần_i just be a person poor, e_i day-day wandered all over hamlet this different e_i beg eat 'Bần was a poor fellow [who] day after day wandered about from one place to another begging for food.'

Sources and Conclusion

There are several worthy examples of grammars and dictionaries of Vietnamese. For the English speaker, there is Nguyễn Đình Hoà (1997) grammar, which is richly exemplified, but reflects the spelling and sometimes the usage before 1975. Thompson (1965, 1987), written in the 1960s, also has a lot of examples, but employs a structuralist model of grammar that might be difficult for some to understand. Some of the examples are also no longer acceptable to contemporary speakers. Cao Xuân Hạo (2003) in an 800-page collection of his essays has discussed Vietnamese from the phonological, grammatical, and semantic perspectives. His bibliography includes many important scholars from the U.S. western Europe, and Russia, such as Bloomfield, Bybee, Chao Yuen Ren, Chomsky, Ducrot, McCawley *et al.*, and a glossary of linguistic terms with English definitions. Notable dictionaries are those by Nguyễn Đình Hoà in many editions, Bùi Phụng (1995) in many editions, and Viện Ngôn Ngữ Học (2000), the model for the contemporary language and the standard-setting dictionary from the Linguistics Institute of Vietnam.

Despite the obvious influences of contact, Vietnamese shows a surprising number of unique features (e.g., tense markers for past and future, some right- and some left-headed typological features), arguably the richest set of pronouns in East and Southeast Asia as well as properties typical of the linguistic area (e.g., a fully developed tone-voice quality sound system, a sharply reduced coda inventory, four-syllable elaborate expressions, and a numeral classifier system). Vietnamese is thus ultimately not very similar to Mon-Khmer, cf., Haudricourt (1953), and is certainly not similar to Sinitic, but a language perhaps analogous in its position to Modern English in the sense that it too has lost many features found in related languages. Yet, despite borrowing and shift influences, Vietnamese, like English, remains an independent and distinctive language in its own right.

Bibliography

- Alves M (2000). 'A look at north-central Vietnamese.' To appear in the papers from the 12th meeting of the Southeast Linguistics Society. University of Illinois, Urbana.
- Bùi P (1995). *Từ'-điển Việt-Anh* (Vietnamese-English dictionary). Hà Nội: Nhà Xuất Bản Thế Giới.
- Cao X H (2003). *Tiếng Việt mấy vấn đề ngữ' âm ngữ' pháp ngữ' nghĩa*. (Vietnamese – some questions about its phonology, syntax, and semantics). Đà Nẵng: Nhà Xuất Bản Giáo Dục.
- Chao Y-R (1930). 'A system of tone letters.' *Le Maître phonétique* 45, 24–27.
- Đào D A (1996). *Hán-Việt từ'-điển*. (Sino-Vietnamese dictionary). Hồ Chí Minh: Nhà Xuất Bản Thành Phố Hồ Chí Minh.
- De Rhodes A (1651). *Dictionarium annamiticum Lusitanum et Latinum*. Rome: Sacrae Congregationis de Propaganda Fide.
- Emeneau M B (1951). 'Studies in Vietnamese (Annamese) grammar' (vol. 8). University of California Publications in Linguistics. Berkeley and Los Angeles: University of California Press.
- Ferlus M (1979). 'Lexique thavung-français.' *Cahiers de Linguistique, Asie Orientale* 5, 71–94.
- Ferlus M (1982). 'Spirantisation des obstruantes modiales et formation du système consonantique du vietnamien.' *Cahiers de Linguistique, Asie Orientale* 11(1), 83–106.
- Ferlus M (1991). 'Vocalisme du proto-Viet-Muong.' Paper circulated at the twenty-fourth ICSTLL. Chiang Mai University, Oct. 10–11, 1991.
- Ferlus M (1996). 'Langues et peuples viet-muong.' *Mon-Khmer Studies* 26, 7–28.
- Ferlus M (1997). 'Problèmes de la formation du système vocalique du vietnamien.' *Cahiers de Linguistique, Asie Orientale* 26, 1.
- Ferlus M (1998). 'Les systèmes de tons dans les langues viet-muong.' *Diachronica* 15.1, 1–27.
- Gordina M V (1960). *Osnovn'ie voprosy foneticheskogo stroia vietnamskogo iazyka* (Basic questions about the phonetic principles of the Vietnamese language). Leningrad.
- Haudricourt A G (1951). 'Les voyelles brèves du vietnamien.' *Bulletin de la Société de Linguistique de Paris* 48, 1, 90–93.
- Haudricourt A G (1953). 'La place du vietnamien dans les langues austroasiatiques.' *Bulletin de la Société de Linguistique de Paris* 49, 122–128.
- Haudricourt A G (1954). 'De l'origine des tons en vietnamien.' *Journal Asiatique* 242, 80–81.
- Hoàng V H (1991). *Từ' điển yếu tố Hán Việt thông dụng* (Dictionary of common use Sino-Vietnamese forms). Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Jacques R (2002). *Portuguese pioneers of Vietnamese linguistics/pionniers portugais de la linguistique Vietnamienne*. Bangkok: Orchid Press.
- Lê Văn Lý (1948). *Le parler vietnamien*. Paris: Hu'o'ng Anh.
- Lý Toàn Thắng (1999). *Về vai trò của Alexandre de Rhodes đối với sự' chế tác và hoàn chỉnh chữ' quốc ngữ'.* *Giao lưu văn hóa và ngôn ngữ' Việt-Pháp* (On the role of Alexandre de Rhodes in the creation and spread of the quốc ngữ' writing system). Hồ Chí Minh: Nhà Xuất Bản Thành Phố.
- Nguyễn D H (1973). 'An 18th-century Chinese-Vietnamese dictionary: The book of three thousand characters.' Paper presented at the 183rd meeting of the American Oriental Society.
- Nguyễn Đ H (1992). Vietnamese phonology and graphemic borrowings from Chinese: the book of 3000 characters revisited. *Mon-Khmer Studies* 20, 163–182.
- Nguyễn D H (1995). Vietnamese-English dictionary (*Tuttle Language Library*). Tokyo: Tuttle.
- Nguyễn D H (1997). *Vietnamese: Tiếng Việt không son phấn*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Nguyễn H Q (2001). *Ngữ' pháp Tiếng Việt* (A grammar of Vietnamese). Hà Nội: Nhà Xuất Bản Từ' Điển Bách Khoa.
- Nguyễn T C (1995). *Giáo trình lịch sử' ngữ' âm Tiếng Việt* (Textbook on the history of Vietnamese phonology). Hà Nội: Nhà Xuất Bản Giáo Dục.
- Nguyễn V L & Edmondson J A (1997). 'Tones and voice quality in modern northern Vietnamese: instrumental case studies.' *Mon-Khmer Studies* 28, 1–18.
- Thompson L C (1965, 1987). 'A Vietnamese reference grammar.' *Mon-Khmer Studies*, 13–15.
- Viện Ngôn Ngữ' Học (Linguistic Institute of Vietnam) (2000). *Từ'-Điển Tiếng Việt* (Vietnamese dictionary). Đà Nẵng: Nhà Xuất Bản.
- Vũ V K (1996). *Từ'-điển Nôm* (Chu Nom dictionary). Đà Nẵng: Nhà Xuất Bản.
- Yakhontov S E (1991). 'Về sự' phân loại các ngôn ngữ' ở Đông Nam Châu Á' (On categorizing among the languages of SE Asia). *Ngôn Ngữ' 1*.

Relevant Website

www.osh.netnam.vn.

Vurës

C Hyslop, La Trobe University, Bundoora,
VIC, Australia

© 2006 Elsevier Ltd. All rights reserved.

Vurës, the name currently preferred by the speakers of the language, is referred to in Ethnologue as the Vetumboso or Vures dialect of Mosina. Mosina is a distinct language, now nearing extinction, with approximately eight speakers residing in the village of the same name, in the southeast of Vanua Lava Island, the Banks group of islands, northern Vanuatu. Vurës is now the dominant language spoken on the island, with upward of 1000 speakers. It is spoken predominantly in the villages of Vētuboso, Wasaga, and Kērēbētia and in smaller surrounding hamlets in the southwest of Vanua Lava. Like all languages of Vanuatu, Vurës is a member of the Oceanic subgroup of the Austronesian family. Within Vanuatu, it belongs to the Northeast Vanuatu-Banks Islands branch of the Northern Vanuatu linkage.

The languages of Northern Vanuatu tend to be fairly conservative Oceanic languages. Vurës has a number of features that are representative of the subgroup and others that are more distinctive. There are 15 consonants and nine vowels in the phonemic inventory, the number of vowels being quite high relative to other Oceanic languages. Notable features of the consonant inventory that are frequently observed in Oceanic languages are prenasalized voiced stops and a labialized labio-velar stop and nasal.

The language is nominative-accusative, with the grammatical relations subject and object being distinguished purely by AVO/SV word order. There is no marking of the subject and object within the verb phrase, which is unusual for an Oceanic language. Oblique arguments are mainly marked by prepositions and occur at the periphery of the clause. The language is both agglutinative and synthetic; however, compared to other Oceanic languages, there is not a great deal of morphology. There is very little

inflectional morphology; aspect and negative polarity are marked by prefixes on the verb and by some particles. There are no tense distinctions. Derivational morphology is also limited, which is linked to the fact that many words are precategorial. This means that they can occur in their underived form as members of more than one word class; in particular, many roots occur as nouns and verbs. Further, many verbs are ambitransitive.

The marking of possession is a complex area of the grammar, a characteristic common to Oceanic languages. For nouns that are inalienably possessed, such as kinterms, body parts, and intimate possessions, the possessor is marked directly on the noun as a suffix. For other items, the possessor is marked on a relational classifier that indicates the function that the item has for the possessor. In Vurës there are six classifiers that mark the possessed item as food, drink, transport, domesticated plants and animals, clothing, or a general default category.

Complex predicates are commonly expressed in Vurës by verb serialization. A serial verb construction can combine verbs to express varied meanings and functions, such as causatives, abilitatives, directionals, and aspectual functions, and some less transparent functions. For example the verb *gial* 'to lie, pretend' can serialize with another verb to express the meaning of pretending to perform the action of the other verb.

Bibliography

- François A (2001). *Contraintes de structure et liberté dans l'organisation du discours: Une description du Mwotlap, langue Océanienne du Vanuatu*. Ph.D. thesis. Université Paris-IV Sorbonne.
- Lynch J & Crowley T (2001). *Languages of Vanuatu: a new survey and bibliography*. Canberra: Pacific Linguistics.
- Lynch J, Ross M & Crowley T (eds.) (2002). *The Oceanic languages*. *Curzon Language Family Series*. Surrey: Curzon Press.

W

Wa

J Watkins, School of Oriental and African Studies,
London, UK

© 2006 Elsevier Ltd. All rights reserved.

The variety of Wa described here is known as Paraok/ pəɹəɔk/ by the people who speak it, and is widely recognized as a standard form of the language. The scope of the name ‘Wa’ is complex, but it is the most inclusive and current among the many names used to refer to the speakers of Wa languages and also corresponds to the terms 佤 *Wǎ* in Chinese and in “ဝ” /wə/ in Burmese. Wa encompasses a cluster of some 40 dialects belonging to the Waic subgroup of the Palaungic branch of Northern Mon–Khmer languages and is spoken by up to about one million people living between the Salween and Mekhong rivers, an area straddling the border between northeastern Myanmar (Burma) and China’s southwestern Yunnan province. About two-thirds of the Wa-speaking population live on the Burmese side of the border and one-third on the Chinese side; most Wa are bilingual in Burmese or Chinese, respectively.

The first Wa orthography intended for popular use was developed in the 1930s by the Baptist missionary

Vincent M. Young, whose translation of the New Testament was published in the 1930s. Orthographies devised for Wa have used the Roman alphabet, sometimes with certain modifications. Young’s orthography was ambiguous and inconsistent in a number of ways, for instance, by failing to represent the register contrast, final glottal consonants, and voiced/sonorants. It has been improved in recent years, however, by incorporating certain features of the phonologically faithful orthography developed by Chinese anthropologist-linguists in the 1950s, which is rather different in design. The most widely encountered orthographies may be compared in Table 1.

The syllable-initial consonants of Wa are set out in Table 2. Wa makes a 4-way voicing contrast in initial stops at 4 places of articulation, and has a range of breathy-aspirated sonorants. Complex initials are restricted to labial or velar stop + /l/ or /r/; final consonants are restricted to /p t c k/, nasals, /ʔ/, and /h/.

In Wa, as in other Mon–Khmer languages, the vowel inventory (Table 3) is effectively doubled because each vowel can occur in either of two registers, ‘clear’ and ‘breathy,’ analogous to the ‘head’ and

Table 1 Various orthographies for Wa

| | | | | | | | | |
|---------------------------|--------------------------------|-------|------|-------|------|------|------|-------|
| Transcription | laj | pət | ʔʔʔ | kum | hɔc | ɲʲak | maɪʔ | nɔh |
| Bible Wa spelling | Lai | pawt | au, | kuim | hoit | jak | mai | naw? |
| Revised Bible Wa spelling | Lai | pawt | aux, | keem: | hoit | jak | maix | nawh? |
| PRC Wa spelling | Lāi | bōd | ex, | geem | hoig | nqag | maix | noh? |
| Gloss | letter | write | 1sg | then | yet | read | 2sg | 3sg |
| Translation | ‘Have you read my letter yet?’ | | | | | | | |

Table 2 Wa consonants

| | <i>Bilabial</i> | <i>Dental/alveolar</i> | <i>Palatal</i> | <i>Velar</i> | <i>Glottal</i> |
|-----------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|----------------|
| Plosive | p p ^h b b ^h | t t ^h d d ^h | ʈ ʈ ^h ɖ ɖ ^h | k k ^h ɡ ɡ ^h | ʔ |
| Nasal | m m ^h | n n ^h | ɲ ɲ ^h | ŋ ŋ ^h | |
| Fricative | v v ^h | s | | | h |
| Approximant | | r r ^h | y y ^h | | |
| Lateral approx. | | l l ^h | | | |

‘chest’ registers of Cambodian (Central Khmer). The register contrast in Wa, as in Mon–Khmer generally, has a complex of phonetic correlates, including fundamental frequency, vowel quality, phonation type, and vowel duration. The blend of these in any individual speaker’s production of the register complex may vary. The register contrast cooccurs with final laryngeal consonants, as illustrated by the set of six words in **Table 4**.

The Wa system of personal pronouns (**Table 5**) retains dual number and contrasts inclusive and exclusive second-person pronouns.

In Wa, the Mon–Khmer morphological prefixation system has all but disappeared, leaving only a few nonproductive semisyllabic prefixes – of which by far the most common is /sə/ – and sound alternations that cover a broad, ill-defined range of functions, illustrated in **Table 6**.

Two characteristics of Wa syntax are that modifiers follow what they modify, as do relative clauses: thus ‘the letter that I wrote,’ from **Table 1**, is expressed as *laj pət ʔɿʔ* ‘letter [write I].’ Secondly, VERB SUBJECT OBJECT word order is commonly observed: ‘you read it’ may be translated *ʔj^hak maiʔ nɔh* ‘read you it,’ though

SUBJECT OBJECT VERB word order is also possible. Two sentences in Wa are shown in (1) and (2):

- (1) *hoc kɛʔ tin yɿh tiʔ*
come 2PL here do what?
What did the two of them come here for?
- (2) *hoc tiʔ sɔk puʔ ʔɿʔ*
come SUBORD search younger. 1SG
brother
They came to see my younger brother.

Wa is an isolating language, adept in the formation of periphrastic compounds. There has been extensive borrowing from Shan, and locally from Chinese and Burmese in areas where those languages are spoken. Loaned vocabulary is frequently supported by a generic Wa word, as in **Table 7**.

Functional literacy in Wa is very low. Wa speakers are much more likely to be literate in the national language of the country in which they live, although some villages may organize grassroots schooling, typically undertaken in a Christian context, and some schools in Wa-speaking China provide, in theory, five years of Wa-language education. Government schools in Burma must use only Burmese.

Table 3 Wa vowels

| | Monophthongs | | | Polyphthongs | | | |
|-----------|--------------|---|---|--------------|-----|----|-----|
| Close | i | u | u | iu | iau | ui | iau |
| Mid-close | e | ɤ | o | ia | uai | ua | uai |
| Mid-open | ɛ | a | ɔ | ei | oi | ou | |
| Open | | a | | ai | au | au | |

Table 6 Vestiges of Mon–Khmer morphological affixation in Wa

| | | | | | |
|--------------|---|-----------------|------------|---|----------------------|
| <i>liah</i> | > | <i>ʔgliah</i> | six | > | sixty |
| <i>laŋ</i> | > | <i>ʔglaŋ</i> | long | > | length |
| <i>rauʔ</i> | > | <i>ʔgrauʔ</i> | deep | > | depth |
| <i>pu</i> | > | <i>ʔmbu</i> | thick | > | thickness |
| <i>tiŋ</i> | > | <i>ʔdiŋ</i> | big | > | size |
| <i>kiap</i> | > | <i>sə.ʔgiap</i> | pinch (v.) | > | clip (n.) |
| <i>ʔdaiʔ</i> | > | <i>sə.ʔdaiʔ</i> | eight | > | no change in meaning |
| <i>ʔgauʔ</i> | > | <i>sə.ʔgauʔ</i> | happy | > | no change in meaning |

Table 4 The Register contrast and laryngeal final consonants

| | Clear register | Breathy register |
|---------------|---------------------|-------------------------|
| Open syllable | <i>tɛ</i> ‘sweet’ | <i>tɛ̤</i> ‘peach’ |
| Final h | <i>tɛh</i> ‘lessen’ | <i>tɛ̤h</i> ‘turn over’ |
| Final ʔ | <i>tɛʔ</i> ‘land’ | <i>tɛ̤ʔ</i> ‘wager’ |

Table 7 Two Wa loanwords

| | | | |
|-------------|--|---------------|-------------------------------|
| <i>pliʔ</i> | <i>mak.mɯŋ</i> | <i>ŋɛʔ</i> | <i>teau.suʔ</i> |
| fruit | mango | house | classroom |
| Wa | Shan: <i>မင်းခူး</i> ; <i>mak.mu:ŋ</i> | Wa | Chinese: 教室 <i>jiàoshì</i> |
| = ‘mango’ | | = ‘classroom’ | |

Table 5 Wa pronouns

| Person | Sing. | Dual | Plural |
|-------------|------------------------|-----------------------|-----------------------------------|
| 1st (incl.) | <i>ʔɿʔ</i> (I) | <i>ʔaʔ</i> you and me | <i>ʔeiʔ</i> we (including you) |
| 1st (excl.) | – | <i>yeʔ</i> he and I | <i>yiʔ</i> we (not including you) |
| 2nd | <i>maiʔ</i> you (sg.) | <i>paʔ</i> you two | <i>peiʔ</i> you (pl.) |
| 3rd | <i>nɔh</i> he, she, it | <i>keʔ</i> they two | <i>kiʔ</i> they |

Bibliography

- Diffloth G (1980). 'The Wa languages.' *Linguistics of the Tibeto-Burman Area* 5, 2.
- Diffloth G (1991). 'Palaungic vowels in Mon-Khmer perspective.' In Davidson J H C S (ed.) *Austroasiatic languages: essays in honour of H. L. Shorto*. London: School of Oriental and African Studies.
- Drage G (1907). *A few notes on Wa*. Rangoon: Superintendent of Government Printing.
- Henderson E J A (1952). 'The main features of Cambodian pronunciation.' *Bulletin of the School of Oriental and African Studies* 14, 149–174.
- Ladefoged P (1983). 'The linguistic use of different phonation types.' In Bless D & Abbs J (eds.) *Vocal fold physiology: contemporary research and clinical issues*. San Diego: College Hill Press. 351–360.
- Scott Sir J G (1896). 'The wild Wa: a head-hunting race.' *Asiatic Quarterly Review* 1–2, 138–152.
- Shorto H L (1963). 'The structural patterns of the northern Mon-Khmer languages.' In Shorto H L (ed.) *Linguistic comparison in Southeast Asia and the Pacific*. London: School of Oriental and African Studies. 45–61.

- Svantesson J-O (1983). 'The phonetic correlates of register in Wa.' In Cohen & van den Broeck (eds.) *Abstracts of the 10th International Congress of Phonetic Sciences*. Dordrecht: Foris.
- Svantesson J-O, Wáng Jīngliú 王敬骝 & Chén Xiāngmù 陈相目 (1981). 'MonKhmer languages in Yúnnán.' *Asie du Sud-Est et Monde Insulindien* 12(1–2), 91–100.
- Watkins J (2002). *The phonetics of Wa: experimental phonetics, phonology, orthography and sociolinguistics*. Canberra: Pacific Linguistics, Australian National University. [Pacific Linguistics 531.]
- Zhōu Zhízhì 周直志 & Yán Qíxiāng 颜其香 (1984). *Wǎyǔ jiǎnzhi 佤语简志*. [Handbook of Wa]. Běijīng: Mǐnzú chūbǎnshè.
- Yán Qíxiāng 颜其香, Zhōu Zhízhì 周直志, Lǐ Dàoyǒng 李道勇, Wáng Jīngliú 王敬骝, Zhāo Míng 赵明, Táo Wén-jūn 陶文均 & Tián Kāizhèng 田开政 (eds.) (1981). *Pug lāi 'Cix ding Yīe si ndong Lāi Vax mai Lāi Hox'/Wǎ-Hàn jiǎnmíng cídiǎn 佤汉简明词典* [A concise Wa-Chinese dictionary]. Kūnmíng: Yúnnán mǐnzú chūbǎnshè.
- Zhāo Yánshè 赵岩社 (1998). *佤语语法 Wǎyǔ yǔfǎ* [Wa grammar]. Kūnmíng: Yúnnán mǐnzú chūbǎnshè.

Wakashan

J Stonham, University of Newcastle, Newcastle upon Tyne, UK

© 2006 Elsevier Ltd. All rights reserved.

Pilling (1894) attributed the term 'Wakashan' to Captain James Cook's observations in Nootka Sound in 1778 wherein he stated: "I would call them Wakashians, from the word *wakash*, which was very frequently in their mouths," (Cook, 1821: 308). Gallatin (in Pilling, 1894) first used the term to designate the Wakashan family and Boas (1890) solidified the boundaries of the family.

The Wakashan family is spread over Vancouver Island and the adjoining areas of the mainland, including northwestern Washington state and the central British Columbia coast.

Languages

The languages of the Wakashan family may be divided into two major subgroups, a northern and a southern group. The Northern group consists of four main languages: Haisla, Oowekyala and Heiltsuk (mutually intelligible), and Kwakwaka (Kwakiutl), located on northeastern Vancouver Island and adjacent parts of the mainland north as far as Kitimat, British Columbia. Within the Northern branch,

Kwakwaka is the best documented, with an early grammar by Hall (1888) and another by Boas (1900). The most northerly language in the family is Haisla, with at least two dialects: Henaksiala and Haisla proper. Oowekyala is spoken by only a handful of people in the area of Rivers Inlet and is to a large extent mutually intelligible with Heiltsuk. Heiltsuk also has two dialects, spoken in Bella Bella and Klemtu, and there is some early work on it by Boas.

The Southern branch of Wakashan consists of three main languages: Nuuchahnulth (Nootka), Ditidaht (Nitinat), and Makah, located on the west coast of Vancouver Island and the tip of northwestern Washington state. Within each language there are various, mutually intelligible dialects. Nuuchahnulth is the most widely spoken of this group, constituting a chain of dialects ranging the length of the west coast of Vancouver Island from Brookes Peninsula to Barkley Sound. Ditidaht constitutes the southernmost Wakashan language on Vancouver Island. It has a close relationship with both the more northerly Nuuchahnulth and the more southerly Makah, which is the only Wakashan language spoken outside of Canada, in Washington state.

Recent estimates of the number of speakers of Wakashan languages vary from approximately 600 to 1200 speakers for all languages (Statistics Canada,

Table 1 Proto-Wakashan phoneme inventory

| | | | | | | | | |
|----|----|----------------|----------------|----|-----------------|---|-----------------|---|
| p | t | χ ^a | c ^b | k | k ^ω | q | q ^ω | |
| b | d | λ | dz | g | g ^ω | ɠ | ɠ ^ω | |
| ḑ | t' | χ ^a | č ^b | ḑ | ḑ ^ω | č | č ^ω | ʔ |
| | | ʈ | s | x | x ^{ωd} | x | x ^{ωd} | h |
| m | n | l | y ^c | ɣ | w | | | |
| m̃ | ñ | l' | ý ^c | ɣ' | w̃ | | | |
| | | | i(:) | | u(:) | | | |
| | | | a(:) | | | | | |

^aIn IPA, /tʰ/ and /tʰ'/. ^bIn IPA, /ts/ and /ts'/. ^cIn IPA, /j/ and /j'/. ^dIn IPA, /ɣ/ and /ɣ^ω/.

2003; Cook and Howe, 2004). (By comparison, there are estimated to be 1185 speakers of Gaelic languages in Canada.) Geographically, the Wakashan languages are adjacent to several other language families, including Athabaskan in the north, Chemakuan in the south, and Salish throughout the area.

Proto-Wakashan

Jacobsen (1979b) located the original home of the family on Vancouver Island or adjacent parts of the mainland, from which it has spread north and south. Swadesh (1953) proposed the Proto-Wakashan phoneme inventory shown in **Table 1**. Sapir (1921) suggested that Wakashan constituted one member, along with Salish and Chemakuan, of a Mosan subgroup, which, together with Kutenai and Algonquian-Ritwan, made up a larger superfamily. Swadesh (1962) noted similarities between Wakashan and Eskimo-Aleut, but little has been done to further any of this research recently (see **Areal Linguistics**). For a summary of work on the comparative-historical study of Wakashan, see Jacobsen (1979b).

Phonology

The phonological inventory of all Wakashan languages consists of a large number of consonant phonemes and a relatively small number of vowel phonemes, usually with a vowel length distinction. In the Northern group there is a three-way distinction among obstruents, involving lax, glottalized, and voiced stops, whereas in the Southern branch there is only a lax versus glottalized distinction, with voiced stop reflexes of the original nasal phonemes appearing in Ditidaht and Makah.

Northern Wakashan has a long/short opposition in the vowel system, whereas the Southern group displays a three-way phonemic length distinction that is realized as a two-way contrast on the surface. The phonemic distinction is due to a third category of

'variable-length' vowels that are long in the first two syllables of the word but short elsewhere, as in *-naʔk* 'have' in *tučnaak* 'have a wife' versus *t'anianak* 'have a child.' Within Northern Wakashan, Oowekyala is purported to have glottalized vowels, which may appear only in the first syllable of a word. Howe (2000) provided minimal pairs such as *ma'təla* 'two people working together' versus *matəla* 'swimming' and *ʔa's* 'animal fat, oil, grease, blubber' vs. *ʔas* 'far out at sea or seaward.'

In Northern Wakashan, the domain of primary stress assignment is the entire word, with stress assigned to the first heavy syllable or to the final vowel if no heavy syllable is encountered. In Southern Wakashan, the domain of primary stress is the first two syllables, with weight contributing to the placement. It should be noted that, although most Wakashan languages employ stress assignment, Kortland (1975) observed that Heiltsuk makes tonal distinctions instead.

Syllable structure is similar for all the languages, allowing complex codas but simple onsets. Boas (1947) stated for Kwakwaka that "consonantic clusters do not occur in initial position. Monosyllabic stems are of the types CVC, CVCC, CVVC, CVVCC." Southern Wakashan likewise involves an obligatorily filled onset (one and only one consonant) and potentially complex codas with up to three, or even four, consonants. Oowekyala appears to exhibit the most extreme cases of consonant clusters, according to Howe (2000).

The processes of glottalization ('hardening') and lenition ('softening') are quite unique in Wakashan and are invariably triggered by the attachment of a suffix to a base ending in a potential candidate for the change. In Southern Wakashan, glottalization affects both obstruents and sonorants, changing stops and affricates to their glottalized counterparts, fricatives to laryngealized glides, and sonorants to their laryngealized counterparts. Lenition, which affects only fricatives in Southern Wakashan, converts them to either /y/ or /w/, depending on whether they are labialized or not. These processes are more complex in Northern Wakashan. Boas (1947) provided the examples (somewhat modified for this presentation) from Kwakwaka (**Table 2**).

One final phonological process worth noting is vowel epenthesis in Makah. In this language (and to some extent in the neighboring Ditidaht), there is a co-occurrence restriction against a voiced or glottalized consonant appearing in the onset of the second syllable when the coda of the first syllable is filled. The resulting cluster is broken up by inserting a lengthened copy of the vowel of the first syllable between the two consonants. Compare the Nuuchahnulth forms, *čaqmis* 'tree bark' and *čusyak* 'shovel'

Table 2 Kwakwala lenition and glottalization^a

| Base | | Lenition | | Glottalization | |
|------|---------------|---------------------|-----------------------|----------------|-------------------|
| eep | 'to pinch' | eeb.ayu | 'dice' | eep.id | 'begin to pinch' |
| wat | 'to lead' | wad.ək ^o | 'led' | wat'.eeneʔ | 'act of leading' |
| pəs | 'to flatten' | pəy.aayu | 'means of flattening' | pəapəc.a | 'try to flatten' |
| məx | 'to strike' | mən.ace | 'drum' | maamən.a | 'ready to strike' |
| čuut | 'to be black' | čuul.atu | 'black-eared' | čuul'.əmyu | 'black-cheeked' |

^a/./ indicates morpheme boundary.

Source: Boas (1947).

with the following Makah examples (from Davidson, 2002).

- (1a) čaqaabis
 čaq-bis
 bark-collectivity.of
 'tree bark'
- (1b) čusuuyak
 čus-yak
 dig-thing.for
 'shovel'

Other common phonological processes include labialization of back consonants, delabialization, and various forms of coalescence, syncope, and epenthesis.

Morphology

These languages are all highly polysynthetic, with complex morphophonemics, and rely heavily on suffixation. Reduplication is the only productive morphological operation that results in a preposed element. There are large numbers of both derivational and inflectional suffixes, but only one root may appear in each word, resulting in an absence of lexical compounding of the usual sort.

Lexical suffixes in Wakashan, unlike Salish, run the gamut of possibilities, including verbal, nominal, adjectival, locative, and adverbial functions. Perhaps the most interesting are the verbal morphemes, which interact with arguments within the sentence, resulting in their combination into a complex predicate, as illustrated in Table 3.

The last example in Table 3 illustrates another property of Wakashan suffixes: the ability of the suffix to trigger various effects on the stem to which they attach, including reduplication and vowel lengthening. The following examples from Makah illustrate these triggers (adapted from Davidson, 2002, where [L] indicates lengthening of the first vowel).

- (2a) hihitaxsʔiq
 hita-xsa[R]-ʔiq
 empty.root-in.bushes-DET
 'in the bushes'

Table 3 Verbal suffixes in Kwakwala

| Verb suffix | Words |
|------------------------------|--|
| -(g)ila 'to make' | χeenagila 'to make oil' χaawayuq ^w ila 'to make a salmon weir' |
| -amāla 'to quarrel about' | suupamāla 'to quarrel over an axe' kəl ^w amāla 'to quarrel over a digging stick' |
| -(g)əʔala[R] 'to wear' | haahaxagəʔala 'to wear a shirt' |

Source: Boas (1947). [R] indicates reduplication.

Table 4 Uses of reduplication in Kwakwala

| | | |
|-----------------------------|---|--|
| Distributive/ plural | g ^y uk 'house' | g ^y ig ^y uk ^w 'houses' |
| Diminutive | nāla 'day' t'eesəm 'stone' bəg ^w 'man' | nānāla 'days' t'at'edzəm 'small stone' baabagəm 'boy' |
| Repetitive | meexa 'to sleep' hanʔa 'to shoot' | meemexa 'to sleep repeatedly' hanthanʔa 'to shoot repeatedly' |
| With derivational suffix | -(g)əʔala[R] 'to wear' | haahaxagəʔala 'to wear a shirt' |

Source: Boas (1947).

- (2b) yuuxłapaal
 yuxł-əpi[L]-' aχ-'i
 float-in air-TEMP=3.SING.IND
 'snow is blowing in the air'
- k^wisi
 k^wisi
 snow

Reduplication is a highly productive process used to indicate a number of distinct grammatical categories. As shown in Table 4, it is employed in derivation, aspect, and inflection as a marker of the distributive or plural, diminutive, and repetition or iteration and as a concomitant of certain derivational suffixes.

Wakashan also employs a set of classifiers that categorize nouns for the purpose of enumerating

Table 5 Kwakwala classifiers

| Number | | /-uk ^o / 'person' | /-sgem/ 'round object' | /-x̩.a/ 'dish, spoon' |
|--------|---------|------------------------------|------------------------|---------------------------|
| nəm | 'one' | nəmuk ^o | nəmsgem | nəmæx̩.a |
| m̩al | 'two' | m̩al'uk ^o | m̩al'tsəm | m̩alæx̩.a |
| yud | 'three' | yuduk ^o | yuduxsəm | yudæx̩ ^o æx̩.a |

Source: Boas (1947).

and qualifying, as well as in a pronominal function, as shown by the examples from Kwakwala in Table 5.

The various combinations of suffixes may result in rather long and complex words, as demonstrated by the examples from Nuuchahnulth in (3) (from Sapir and Swadesh, 1939. NOW 'contemporaneous,' SW 'switch reference,' CLS 'classifier').

- (3a) ʔuuq̣x̩n̩uk^w aʔatquuʔ
 ʔu -ʔaq̣x̩ -n̩uk^w -ʔaʔ -ʔat -quuʔ
 REF -inside -inhand -NOW -SW -3S.CND
 'if one is holding it' (Source: Sapir and Swadesh, 1939)

- (3b) ʔaʔaʔaʔqim̩t̩im̩iʔm̩in̩h̩ʔaaq̣x̩eʔicu
 DUP- DUP- ʔaʔ -qim̩t̩ -h̩ta -maʔ
 REP- SUFF two -CLS -onfoot[R] -move
 -ʔit -m̩in̩h̩ -ʔaaq̣x̩ -(m)eʔicu
 -onfloor -PL -INTENT-2PL.IND
 'You will carry two dollars on your feet' (Source: Sapir and Swadesh, 1955)

Syntax

Basic word order for Wakashan languages is, in general, head-initial, with VSO being the most common order. The degree to which individual languages allow the transposition of subjects and objects is one area of variation within the family. There is no case marking on nominals, but in some contexts complex prepositions may be used to indicate the grammatical role of arguments, as in the following example from Ditidaht (adapted from Klokeid, 1978).

- (4) ʔuq^wsiʔa ʔuʔ^w John ʔuuyuq^w Bill
 hit NOM John ACC Bill
 'John hit Bill'

Within the nominal phrase, quantifiers precede adjectives which, in turn, precede the head noun, as exemplified by the following examples from Nuuchahnulth (Rose, 1981). Relative clauses follow the head noun, as in (5b).

- (5a) ʔi̩h̩ saya n̩is̩m̩a
 very distant land
 'a really distant land'
- (5b) ha x̩utaay [yaaq̣h̩w̩aʔ naq Bill]
 DET knife that.used Bill
 'the knife that Bill used'

There are well-developed person-number inflectional paradigms that typically appear after the first position in the sentence in clitic-like fashion. Boas (1900: 715) remarked on "... the tendency of adverbs and auxiliary verbs to take the subjective ending of the verb, while the object remains connected with the verb itself. *k̩eʔsən d̩uuquaʔ* not-I see-him, shows the characteristic arrangement of sentences of this kind."

Possession associated with the arguments of the clause is sometimes marked on the predicate, as shown in the examples in (6) from Kwakwala (Boas, 1900).

- (6a) n̩eekən ɡənəm
 say.my wife
 'my wife said'
- (6b) n̩eekexən ɡənəm
 say.he.my wife
 'he said to my wife'
- (6c) n̩eekexees ɡənəm
 say.he.his wife
 'he said to his (own) wife' (Boas, 1900)

Tense markers exhibit the special Wakashan characteristic of appearing on both nouns and verbs, leading to the common conclusion that there are no category distinctions in these languages (however, cf. Jacobsen, 1979a).

A form of syntactic compounding exists, at least in some members of the family, as illustrated by the following examples from Nuuchahnulth.

- (7a) ʔi̩h̩ʔi̩ [yaʔmuut ʔaq̣mis]
 the big bladder oil
 'the large oil bladder' (Sapir and Swadesh, 1939)
- (7b) [muunaa n̩i̩q̣n̩i̩q̣aʔak]
 machine sew -tool
 'a sewing machine' (Sapir and Swadesh, 1955)

Further Reading

For further information on Wakashan, the reader is referred to the references appended to this article, in particular the discussion in Boas (1947), Davidson (2002), Howe (2000), Jacobsen (1979a, 1979b), Lincoln and Rath (1980, 1986), and Stonham (1999, 2004).

Bibliography

- Boas F (1890). 'The Nootka.' *Report of the 60th meeting of the British Association for the Advancement of Science* 60, 582–604.
- Boas F (1900). 'Sketch of the Kwakiutl language.' *American Anthropologist* 2, 708–721.
- Boas F (1947). 'Kwakiutl grammar with a glossary of the suffixes.' *Transactions of the American Philosophical Society* 37(3), 202–377.
- Cook E-D & Howe D (2004). 'Aboriginal languages of Canada.' In O'Grady W & Archibald J (eds.) *Contemporary linguistic analysis*, (5th edn.). Toronto, Canada: Addison Wesley Longman. 294–309.
- Cook J (1821). *The three voyages of Captain James Cook round the world*. Complete in Seven Volumes. Vol. 6. London: Longman, Hurst, Rees, Orme, and Brown.
- Davidson M (2002). 'Studies in Southern Wakashan (Nootkan) grammar.' Ph.D. diss., State University of New York at Buffalo.
- Hall A J (1888). 'A grammar of the Kwagiutl language.' *Transactions of the Royal Society of Canada* 6(2)I.
- Howe D M (2000). 'Oowekyala segmental phonology.' Ph.D. diss., University of British Columbia.
- Jacobsen W H Jr (1979a). 'Noun and verb in Nootkan.' In Efrat B S (ed.) *British Columbia Provincial Museum Heritage Record 4: The Victoria Conference on North-western Languages*. Victoria, Canada: British Columbia Provincial Museum. 83–155.
- Jacobsen W H Jr (1979b). 'Wakashan comparative studies.' In Campbell L & Mithun M (eds.) *The languages of Native America*. Austin, TX: University of Texas Press. 766–791.
- Klokeid T J (1978). 'Surface structure constraints and Niti-naht enclitics.' In Cook E-D & Kaye J (eds.) *Linguistic studies of native Canada*. Vancouver, Canada: University of British Columbia Press. 157–176.
- Kortlandt F H H (1975). 'Tones in Wakashan.' *Linguistics* 146, 31–34.
- Lincoln N J & Rath J C (1980). *North Wakashan Comparative Root List*. National Museum of Man Mercury Series, Canadian Ethnology Service Paper 68. Ottawa, Canada: National Museums of Canada.
- Lincoln N J & Rath J C (1986). *Phonology, dictionary and listing of roots and lexical derivatives of the Haisla language of Kitlope and Kitimaat*, B.C. Canadian Museum of Civilization Mercury Series, Canadian Ethnology Service Paper 103. Ottawa Canada: National Museums of Canada.
- Pilling J C (1894). *Bibliography of the Wakashan languages*. Washington, D.C: Government Printing Office.
- Rose S M (1981). 'Kyuquot grammar.' Ph.D. diss., University of Victoria.
- Sapir E (1921). 'A bird's-eye view of American languages north of Mexico.' *Science* 54, 408.
- Sapir E & Swadesh M (1939). *Nootka texts, tales and ethnological narratives, with grammatical notes and lexical materials*. Philadelphia and Baltimore, MD: Linguistic Society of America.
- Sapir E & Swadesh M (1955). *Native accounts of Nootka ethnography*. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore and Linguistics.
- Statistics Canada (2003). *Aboriginal peoples of Canada: a demographic profile*. (January 21, 2003 report on the 2001 Census.) <http://www.statcan.ca:8096/bsolc/english/bs.c?catno=96F0030X2001007>.
- Stonham J (1999). *Aspects of Tsishaaht Nootka phonetics and phonology*. Munich: LINCOM Europa.
- Stonham J (2004). *Linguistic theory and complex words: Nuuchahnulth word formation*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Swadesh M (1953). 'Mosan I: a problem of remote common origin.' *International Journal of American Linguistics* 19, 26–44.
- Swadesh M (1962). 'Linguistic relations across Bering Strait.' *American Anthropologist* 64, 1262–1291.

Wambaya

R Nordlinger, University of Melbourne, Melbourne, Australia

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Wambaya is a non-Pama–Nyungan language of northern Australia (of the Mirndi group), originally spoken in the Barkly Tablelands region of the Northern Territory. The Wambaya people suffered greatly from the invasion of their land, their subsequent removal from their traditional country, and the dispersal of their

community. As a result of these and other factors, the Wambaya language has now almost disappeared. There are only a handful of (semi-)speakers remaining, most of whom live in the towns of Tennant Creek, Elliott, and Borroloola. Wambaya is closely related to two further dialects—Gudanji and Binbinka. Gudanji is in much the same state as Wambaya with only a handful of (semifluent) speakers left; and there are no longer any remaining speakers of Binbinka.

Like all Australian languages, Wambaya was not traditionally written, and so the earliest records we have of the language are those collected by white researchers since the early 20th century. Some lexical

items were recorded by Mathews (1900, 1908) and by Spencer and Gillen (1904), largely concerning the kinship system. The first detailed grammatical information for Wambaya is found in the field notes recorded by Ken Hale in 1959 (Hale, 1959, 1960) and is continued in Neil Chadwick's work on the whole Barkly language group (including also Jingulu [Djingili] and Ngarnka [Ngarndji]) dating from the 1970s (Chadwick, 1978, 1979, 1984, 1997). My own fieldwork on the language began in 1991 and has so far resulted in the publication of a grammar of Wambaya (Nordlinger, 1998a), the development of a dictionary and learner's guide (Nordlinger, 1998b, 1998c), and a number of articles (Nordlinger, 1995, 2001; Nordlinger and Bresnan, 1996; Green and Nordlinger, 2004). The grammatical features discussed in this brief article are all discussed and exemplified in greater detail in Nordlinger (1998a).

Phonology

Phonologically, Wambaya is a typical Australian language, with five places of articulation for stops, including two apical series (apico-alveolar and retroflex) and one laminal series. There is no voicing contrast. There is a nasal corresponding to each stop articulation, and three laterals (in the nonperipheral places of articulation). There are three vowel phonemes, and no productive length distinction. The phoneme inventory and the orthographic symbols corresponding to each phoneme are provided in Table 1.

All Wambaya words are minimally disyllabic and virtually always vowel-final. (The one exception is the auxiliary, see below, which can end in a consonant if it contains one of three nasal-final affixes: *-any* 'direction away, past tense,' *-amany* 'direction towards, past tense,' or *-n* 'progressive aspect'.) Primary stress is generally on the first syllable of the word. Words can begin with a vowel (as in *alaji* 'boy'), or a consonant (*daguma* 'hit,' *juwa* 'man,' *ngajbi* 'see'). There are no words beginning with the consonants *r* [ɹ], *rr* [ɹɹ], *ly* [ɬ] and further, as is common among Australian languages, the distinction between the two

apical series is neutralized in initial position. Initial apicals are all represented orthographically as apico-alveolars (*d*, *n*, *l*). Biconsonantal clusters are common, but only word-medially. Such clusters usually contain an apical or laminal consonant followed by a labial or dorsal consonant (e.g., *anmurru* 'cuddle,' *bardgu* 'fall,' *marrgulu* 'egg,' *ngajbi* 'see,' *manganyma* 'tucker, nonmeat food'), but other combinations are possible also (*bungmaji* 'old man,' *wugbardi* 'cook'). There is one triconsonantal cluster *rrgb* (*lurrgbanyi* 'grab,' *gurrgbarra* 'stare').

A brief discourse in the language written in the conventional orthography is provided in (1). This example illustrates many of the grammatical properties to be discussed below.

- (1) Yarru ngurr-any gurdi-nmanji ngaj-barda.
 go.NF 1PL.INC-PST.AWAY bush.IV.OBL-ALL see-INF
 Gannga ngurr-amany bangarnigadi.
 return.NF 1PL.INC-PST.TWDS this way
 Gurijba gi-n mirra
 good.IV.NOM 3SG.S.PRES-PROG sit.NF
 ngarrga maga.
 my.IV.NOM house.IV.NOM
 'We went to the bush to have a look (at my house).
 Then we came back this way. My house is fine.'

Morphology

Wambaya, being one of the southernmost non-Pama-Nyungan languages, is typologically atypical in having lost virtually all prefixing morphology (see Nordlinger, 1998a; Green and Nordlinger, 2004; Harvey *et al.*, to appear for discussion). All productive morphology is suffixing. Like many other Australian languages, Wambaya has extensive case and agreement morphology – all elements of an NP must show concord in gender, number, and case – and 'free' (i.e., pragmatically determined) word order.

The two open word classes are verbs and nominals. Concepts translated into European languages as adjectives are split across these two classes. For example, *bulyingi* 'small,' *bugayi* 'big,' *gurijbi* 'good,' and *bagijbi* 'bad' are nominals, while *baliji* 'be

Table 1 Wambaya phonemes

| Consonants | Bilab. | Apico-alv. | Apico-postalv. (retroflex) | Lamino-palatal | Velar |
|------------|------------------|-----------------|----------------------------|----------------|--------|
| Stop | b (b) | d (d) | ɖ (rd) | ʃ (j) | g (g) |
| Nasal | m (m) | n (n) | ɳ (rn) | ɲ (ny) | ŋ (ng) |
| Lateral | | l (l) | ɭ (rl) | ʎ (ly) | |
| Tap/Trill | | ɾ/r (rr) | | | |
| Semivowel | w (w) | | ɹ (r) | j (y) | |
| Vowels | i (i) (i: (iii)) | | u (u) | | |
| | | a (a) (a: (aa)) | | | |

hungry,’ *gurda* ‘be sick,’ and *laji* ‘be quiet’ are intransitive verbs. Verbs are characterized by the fact that they must always cooccur with the auxiliary (see below), which carries subject/object agreement information and tense/aspect/mood for the clause. Verbs have only a small amount of inflectional morphology, contrasting future/imperative, and nonfuture (unmarked) forms. This lack of morphology appears to result from the fact that these are historically derived from uninflected coverbs, with the synchronic auxiliary being the original inflected main verb (see Green and Nordlinger, 2004 and below for discussion). There are two inflectional classes for verbs, membership of which is phonologically conditioned. Vowel-final verb stems belong to the J-class of verbs, consonant-final verb stems belong to the Ø-class of verbs (and there are, of course, a dozen or so irregular verbs that don’t follow the pattern of either class). These classes are distinguished by the fact that in the J-class, there is a thematic *-j-* added before any verbal affixes, while there is no thematic consonant in the Ø-class. The two classes also have different nonfuture tense inflections. Examples of the two paradigms are:

- (2) *daguma*- ‘hit’ (J-class): *daguma* (nonfuture),
daguma-j-ba (future/imperative), *daguma-j-*
barda (infinitive)
- (3) *gulug*- ‘sleep’ (Ø-class): *gulug-bi* (nonfuture),
gulug-ba (future/imperative), *gulug-barda*
(infinitive)

In contrast to verbs, nominals have a large amount of morphology. As is relatively common among northern Australian languages, there are four genders – masculine (class I), feminine (class II), vegetable (class III) (including nonmeat food and some body parts), and neuter (class IV) (residue). These genders are marked on nominals and their modifiers by suffixes, and are marked on demonstratives by cognate prefixes (vestiges of the earlier prefixing system, see Green and Nordlinger, 2004 and Harvey *et al.*, to appear for discussion). Example (1) above shows concord with adjectives (*gurijba* ‘good.IV’) and possessive pronouns (*ngarrga* ‘my.IV’). That these forms are truly agreeing with the head nominal *barrawu* ‘house’ can be shown by contrast with the following example in which they are agreeing instead with a class I nominal *janji* ‘dog.’

- (4) yini ngarri janji gurijbi
this.I my.I dog.I good.I
‘this is my good dog’

Nominals are also inflected for number – dual and plural – although since the unmarked nominal can have singular, dual, or plural interpretations, this inflection is optional. Pronouns (a subclass of nominals)

Table 2 Wambaya core case system

| | A | S | O |
|----------|----------|------------|------------|
| Nominals | Ergative | Nominative | Accusative |
| Pronouns | Ergative | Nominative | Accusative |

distinguish singular, dual, and plural numbers and make an inclusive/exclusive distinction in the 1st person nonsingular.

Nominals are obligatorily inflected for case. Wambaya is what has been called a ‘split-ergative’ language. Nominals inflect according to an ergative/absolutive distinction, while pronouns inflect largely on a nominative/accusative pattern. We can, therefore, distinguish three distinct case contrasts, as in Table 2, where A stands for ‘transitive subject,’ S for ‘intransitive subject,’ and O for ‘transitive object.’ In the following table, the shading indicates homophony between forms; thus, nominals have homophonous nominative and accusative forms, while pronouns have homophonous ergative and nominative forms.

As well as the three core cases shown in Table 2, there are nine further cases in Wambaya, including: dative, ablative, allative, comitative, genitive, proprietive, privative, perlative, causal, and originative. The ergative case additionally covers locative and instrumental functions. Gender markers also distinguish case, having one form before cases with zero realizations (i.e., the nominative and the accusative), and another form (called the ‘oblique’ form) before all other cases. For example, *jan-ji* ‘dog-I.NOM’ vs. *jan-yi-ni* ‘dog-I.OBL-ERG’; *guji-nya* ‘mother-II.NOM’ vs. *guji-ga-nka* ‘mother-II.OBL-DAT’; *mangany-ma* ‘food-III.NOM’ vs. *mangany-mi-nka* ‘food-III.OBL-DAT.’

All verb-headed clauses in Wambaya must contain a grammatical auxiliary containing subject and object cross-referencing bound pronouns and clausal tense/aspect/mood information. While word order is grammatically free in Wambaya for the most part, the auxiliary is unusual in having a fixed position: it must always occur in second position in the clause (after the first constituent, which may be a single word or a complex NP). The auxiliary appears to have derived from a fully inflecting verb with pronominal agreement prefixes and tense/aspect/mood suffixes. An original main verb–coverb structure (as is common in other northern Australian languages, see McGregor, 2002), has become an auxiliary-verb structure in Wambaya, with the auxiliary retaining only the grammatical information of the original main verb, and the original coverb now contributing all lexical meaning. Remnants of an original main verb are found in the directional/tense

- Nordlinger R (1998a). *A grammar of Wambaya (Northern Territory, Australia)*. Canberra: Pacific Linguistics.
- Nordlinger R (1998b). *An elementary Wambaya dictionary (with English–Wambaya finder list)*. MS: Australian Institute for Aboriginal and Torres Strait Islander Studies.
- Nordlinger R (1998c). *A learner's guide to basic Wambaya*. MS, Australian Institute for Aboriginal and Torres Strait Islander Studies.

- Nordlinger R (2001). 'Wambaya in Motion.' In Simpson J, Nash D, Laughren M, Austin P & Alpher B (eds.) *Forty years on*. Canberra: Pacific Linguistics. 401–413.
- Nordlinger R & Bresnan J (1996). 'Nonconfigurational tense in Wambaya.' *Proceedings of the LFG Conference*, Grenoble, August 1996. CSLI Publications.
- Spencer B & Gillen F J (1904). *Northern tribes of central Australia*. London.

Warlpiri

R Hoogenraad, Alice Springs, NT, Australia

© 2006 Elsevier Ltd. All rights reserved.

Warlpiri is a Pama-Nyungan language of the Ngumbin-Yapa subgroup (see McConvell and Laughren, 2004) spoken by some 3000 *Yapa* 'people' (typically Warlpiri people). The Warlpiri heartland is the Tanami Desert to the northwest of Alice Springs in Australia's Northern Territory, but the Warlpiri-speaking population now lives mainly in four communities around the margins of this area: Lajamanu, Nyirrpri, Yurntumu, and Wirliyajarrayi. There are also sizable populations in Katherine, Tennant Creek and Alice Springs, and in Aboriginal communities around the Warlpiri area. Warlpiri is also used over a larger area as a lingua franca – possibly up to 1000 Aboriginal people speak Warlpiri as a second language.

Warlpiri had seven dialects, which are now being reduced to four communalects: Yurntumu/Nyirrpri, Lajamanu, Wirliyajarrayi, and Wakirti Warlpiri (Alekenenge and Tennant Creek).

All the neighboring languages belong to the Pama-Nyungan family: Ngumbin-Yapa subgroup (Warlpiri, Ngardi, Jaru, Nyininy, Gurindji, Mudburra, Warlmanpa), Warumungu, Western Desert group (Pintupi Luritja, Pintupi, Kukatja), Arandic Group (Alyawarr, Kaytetye, Central Anmatyerr). Though there are significant differences between them, they share many of their core structural and sociolinguistic features.

Warlpiri Grammar

Ken Hale was Warlpiri's 'recording angel,' starting in 1959 (see the bibliography in Simpson *et al.*, 2001). There is a learner's guide (Laughren *et al.*, 1996), and good overviews are provided by Nash (1986), Hale *et al.* (1995), Simpson (1991), and the collection of papers in Swartz (1982a).

A Warlpiri encyclopedic dictionary is in preparation: the database currently has over 9500 entries. Hale (1995) is a simple dictionary with nearly 2000 entries, and has an appendix with a concise grammatical inventory (as has Hale *et al.*, 1995).

Phonology

Table 1 shows the consonant phonemes in the standard orthography in use since 1974. The contrast between postalveolar stop *rt* and flap *rd* is only allophonic in eastern dialects: Wirliyajarrayi and Wakirti Warlpiri.

There are three vowels, *i*, *u*, and *a*. High vowels *u* and *i* harmonize with adjacent high vowels across morpheme boundaries: progressive in nominals, *wati-ngki* man-ERG, *kurdu-ngku* child-ERG, but regressive in verbs, *kipi-rni* winnow-PRES, *kupu-rnu* winnow-PAST. The low vowel *a* blocks vowel harmony: *kirlilkirlilpa-rlu* galah-ERG; *yirra-rni* put-PRES, *yirra-rnu* put-PAST. A syllable may have one 'short' vowel or a sequence of two identical vowels, e.g., *ngurrpa* 'ignorant of,' *nguurrpa* 'windpipe.'

The Word

Warlpiri words must have at least two vowels, in the same or sequential syllables, with an initial consonant

Table 1 Warlpiri consonants

| | Peripheral | | Coronal | | |
|----------|------------|-----------|----------------|--------------------------------|----------------|
| | Bilabial | Velar | Lamino-palatal | Apico-postalveolar (retroflex) | Apico-alveolar |
| Stop | <i>p</i> | <i>k</i> | <i>j</i> | <i>rt</i> | <i>t</i> |
| Nasal | <i>m</i> | <i>ng</i> | <i>ny</i> | <i>rn</i> | <i>n</i> |
| Lateral | | | <i>ly</i> | <i>rl</i> | <i>l</i> |
| Flap/tap | | | | <i>rd</i> | <i>rr</i> |
| Glide | <i>w</i> | | <i>y</i> | | <i>r</i> |

and a final vowel and stress on the first syllable. The alveolar/postalveolar distinction (see Table 1) is neutralized word initially. The laminopalatal *ly* only occurs following a vowel.

There is a mora (or vowel) counting rule that determines the choice between the two forms of the locative (*-ngka* vs. *-rla*) and ergative (*-ngku~i* vs. *-rlu~i*), with minimal words hosting the velar allomorph, e.g., *ngurrpa-ngku* vs. *nguurrrpa-rlu* (Hale, 1995: Appendix).

Clauses

Warlpiri has a classical nonconfigurational clause structure, with ‘free’ word order, discontinuous constituents, and a high reliance on zero anaphora (Hale, 1983; Hale *et al.*, 1995; Swartz, 1991). It has verbal and nominal (i.e., verbless) clauses. The nominal clause can be rendered as a verbal clause using one of the stance verbs as copula (see Table 2).

Ngapa ngurruju. Ngapa-ka karri-mi ngurruju.
water good water-AUX vertical.(stand)-PRES good
 ‘The water is good.’

The verbal clause consists of an obligatory verb and associated auxiliary complex, and a number of optional case-marked nominal constituents, as well as adverbial enclitics and particles with clausal or subclausal scope (Laughren, 1982a). Dependent and nonfinite subordinate clause types are discussed in Hale (1976) and Hale *et al.* (1995).

Verbs

There are 170 simple verb stems, which inflect for tense and mood in five conjugations. Past, nonpast

Table 2 Warlpiri stance verbs/copula (verb ‘to be’)

| Neutral | Vertical | Horizontal | Humped |
|-------------------------------|---------------------------|-------------------------|--------------------------------|
| <i>nyinami</i> ‘sit, stay’ | <i>karrimi</i> ‘stand’ | <i>ngunami</i> ‘lie’ | <i>parntarrimi</i> ‘crouch’ |

and irrealis verbs co-occur with aspect auxiliaries (see Table 3), while the imperative, infinitive, future, and presentational forms do not. Derivational morphology creates a nomic or agentive/instrumental noun from verbs (Hale, 1995), and inceptive and iterative verb forms which inflect for tense and mood, e.g., *payirni-njina* ‘go and ask’ vs. *payirni-njina-na* ‘go about asking.’

Verbs are fixed in one of five transitivity patterns specifying syntactic case arrays: intransitive, bi-intransitive, middle, transitive, and bi-transitive (Swartz, 1982b).

The inchoative *-jarri-* and the transitivizer *-ma-* are two very productive verb formatives with nominal stems. They form interrogative verbs: *Nyarrpa-jarrija?* ‘What did (he) do?’ and *Nyarrpa-manu?* ‘What did (he) do to (it)?’

Verbal meanings are further expanded by a large open noninflecting category termed preverb, particularly with the eight monosyllabic verb roots (Nash, 1982): *nyanyi* ‘seeing,’ *purda-nyanyi* ‘hearing,’ *parnti-nyanyi* ‘smell.’ Nominals, e.g., *jarda* ‘sleep, asleep,’ can act as preverbs; *jarda-ngunami* ‘asleep-lying, sleeping.’

Nominals

Nominals make up the largest grammatical class, and include substantives (nouns) and descriptive terms (adjectives), free pronouns, the deictics and determiners, question words, etc. In verbal clauses they are optional. Nominals host number-marking suffixes (see Table 4), followed by a range of syntactic and semantic case suffixes (Hale *et al.*, 1995).

The three syntactic cases are in an ergative system. The ergative, *-ngku~i* or *-rlu~i*, marks the subject (agent) of a transitive verb. The absolutive, *-Ø* (zero), marks the subject of an intransitive verb and the object of a transitive verb. The dative, *-ku~i*, marks indirect object and some oblique functions. There are a number of spatial cases, listed in Table 5, as well as an alienable possessive case. A grammatical

Table 3 The Warlpiri auxiliary complex

| Aspect and tense | | | | | |
|---|---|--|---|---|---|
| -Ø- (i.e., nothing) | | | -ka- | -lpa- | |
| Perfective | | | Imperfective | | |
| Nonpast | Past | Irrealis | Nonpast | Past | Irrealis |
| Immediate probability ngarrirni 'about to tell, may tell' | Completed action ngarrurnu 'has told, told' | Past possibility, probability ngarrikarla 'would/should have told' | Happening; habitually happens -ka ngarrirni 'is telling, tells' | Action in progress in past -lpa ngarrurnu 'was telling' | Possibility, probability -lpa ngarrikarla 'would/should tell' |

Table 4 Warlpiri grammatical number

| | <i>Singular</i> (one) | <i>Dual</i> (a pair) | <i>Paucal</i> (several) | <i>Plural</i> (more) |
|---------------------------------------|--------------------------|-------------------------|----------------------------|-------------------------|
| Nouns, interrogatives, adjectives | -Ø | -jarra | -patu | -Ø |
| Definite deictic and determiners | -Ø | -jarra | -patu | -rra |
| Indefinite determiners | jinta | jirrama | marnkurpa | panu |
| 3rd Person subject pronominal clitics | -Ø | -pala | | -li~lu |
| 3rd Person object pronominal clitics | -Ø | -palangu | | -jana |

Table 5 Warlpiri spatial cases

| | | | | |
|-----------|--------------------------|--------------------------------|---------|----------------------------|
| -ngka~rla | locative 'at, on, in' | -kurra allative 'to, up to' | -ngurlu | elative 'from' |
| -wana | 'along, around' | | -jangka | 'from' origin, cause |

case may suffix to a stem containing a semantic case (see Simpson, 1991).

Auxiliary Complex

The auxiliary complex is in the Wackernagel position, i.e., after the first constituent of the clause. The auxiliary complex comprises:

- an optional finite complementizer (Hale *et al.*, 1995)
- an aspect marker, which in concert with the verb tense specifies the temporal and aspectual meaning of the clause (Table 3)
- up to three pronominal clitics marking subject and nonsubject functions (Hale, 1973, 1982).

There is a systematic mapping of the syntactic case array specified by the verb – in an ergative system – onto the auxiliary pronominal clitics – in a subject-object system – described from different perspectives by Swartz (1982b) and Hale (1982).

The free pronouns and the pronominal clitics mark person and number (Table 4). Third person singular subject and direct object is unmarked. A clause can be very minimal, without any overt nominals or auxiliary morphemes, or it can be expanded with case-marked nominals, etc.

Jarnturnu.
'He trimmed it.'

can be expanded to:

Wati-ngki-Ø-Ø-Ø karli-Ø jarntu-rnu.
man-ERG-PERF-3rd.sing.SUBJ-3rd.sing.OBJ boomerang-
ABS trim-PAST
'The man trimmed the boomerang.'

Expanding this, and changing the participants and aspect to show additional possibilities:

Table 6 Warlpiri spatial location and orientation: deictics

| | <i>Definite</i> | | <i>Indefinite</i> |
|---------|--------------------------|---|---|
| | <i>Location certain</i> | <i>Not visible</i> | <i>Location uncertain</i> |
| Close | nyampu 'this, here' | yalarnimpi 'this one not visible' | mirnimpi 'somewhere here' |
| Nearby | yalumpu 'that, there' | yalarni 'that one not visible' | mirni 'somewhere there' |
| Further | yali 'yon, yonder' | (yalarnimpay?) 'yonder not visible' | mirnimpay 'somewhere yonder' |
| Distant | yinya 'far yonder' | yalarnirra 'that one far off out of sight' | mirnirra 'far away, a long way away' |

Table 7 Warlpiri direction of action relative to speaker, suffixed to the inflected verb

| <i>Centripital</i> | <i>Centrifugal</i> | <i>Perpendicular</i> |
|--------------------------------|--------------------------|----------------------|
| +rni~rnu 'towards (hither)' | +rra 'away (thither)' | +mpa 'across' |

Ngajarra-ku kula-lpa-Ø-jarrangu-rla karli-Ø
jarntu-rnu wati-ngki

1st.excl.dual-DAT NEG.COMP-IMPERF-3rd.sing.SUBJ-

1st.excl.dual-DAT boomerang-ABS trim-PAST man-
ERG

'Us two (excluding you) was not for whom the man was trimming the boomerang.'

Meaning, Context, and Registers

I now discuss briefly some Warlpiri cultural preoccupations – with land, social relationships, and proper behavior – that are reflected in the language.

Tables 5–7 shows some of the ways in which spatial orientation and direction is encoded in the language. For instance, there is a rich set of deictics encoding distance, visibility, and definiteness, and the choice of copula is determined by the perceived orientation of the subject (see Table 2). In addition, there is a system

of absolute three-dimensional spatial reference based on the four compass directions and up and down. The directional suffixes in Table 7 and additional suffixes create a rich system of spatial reference (Laughren, 1978).

Although Warlpiri had no counting system, Table 4 shows that it marks grammatical number on all nominals and references it in the pronominal clitics. There is a particular emphasis on pairing, reflected in the obligatory marking of dual, and the dual indefinite determiner.

There is also a preoccupation with relationships: the kin terminology allows any Yapa to be referenced as a relation, with higher order groupings into patrilineal, matrilineal, and generation moieties. There is also a sociocentric system of eight subsections or 'skin names.'

Pairing shows up again in an extensive set of tri-relational kin terms, which allow any pair of people to be referenced as a triangulation of the relationship between the speaker and each of the pair, and the relationship between the pair (Laughren, 1982b), e.g., *makurnta-rlangu*, the relationship between one's brother-in-law and mother, who stand in the avoidance relationship of mother-in-law/son-in-law to each other (*kurnta* 'shame, proper behavior'). Relationship to land is also registered in the kinship system.

Warlpiri has respect registers, characterized by obliqueness, used when speaking about or to relations and ritual associates in avoidance or respect relationships (Laughren, 2001). *Rdaka-rdaka* 'Hand Talk' is a sign language used when speaking is inappropriate, especially by bereaved widows and mothers in the *jilimi* 'single women's camp' (Kendon, 1988). Baby Talk is used to talk to babies, incrementally building up the phonological, grammatical, and semantic features of Warlpiri for them as they develop (Laughren, 1984).

All these registers are characterized by hyperpolysemy and a systematic reduction in distinctions, giving us an insight into the organization of Warlpiri semantics, grammar, and phonology.

Bibliography

- Hale K L (1973). 'Person marking in Walbiri.' In Anderson S R & Kiparsky P (eds.) *A Festschrift for Morris Halle*. New York: Holt, Rinehart & Winston. 308–344.
Hale K L (1976). 'The adjoined relative clause in Australia.' In Dixon R M W (ed.) *Grammatical categories in*

- Australian languages*. Canberra, New Jersey: Humanities Press. 78–105.
Hale K L (1982). 'Some essential features of Warlpiri verbal clauses.' In Swartz (ed.). 217–315.
Hale K L (1983). 'Warlpiri and the grammar of non-configurational languages.' *Natural Language and Linguistic Theory* 1, 5–47.
Hale K L (1995). *An elementary Warlpiri dictionary (revised edition)*. Alice Springs: IAD Press.
Hale K L, Laughren M & Simpson J (1995). 'Warlpiri.' In Jacobs J, van Stechow A, Sternefeld W & Vennenmann T (eds.) *Syntax—handbook*, vol. 2. Berlin/New York: Walter de Gruyter. 1430–1451.
Kendon A (1988). *Sign language in Aboriginal Australia*. Cambridge: Cambridge University Press.
Laughren M (1978). 'Directional terminology in Warlpiri.' *Launceston Working Papers in Language and Linguistics* 8, 1–16.
Laughren M (1982a). 'A preliminary description of propositional particles in Warlpiri.' In Swartz (ed.). 129–163.
Laughren M (1982b). 'Warlpiri kinship structure.' In Heath J, Merlan F & Rumsey A (eds.) *Oceania Linguistic Monographs N° 24: Languages of kinship in Aboriginal Australia*. Sydney: University of Sydney. 72–85.
Laughren M (1984). 'Warlpiri baby talk.' *Australian Journal of Linguistics* 4, 73–88.
Laughren M (2001). 'What Warlpiri "avoidance" registers do with grammar.' In Simpson *et al.* (eds.). 199–225.
Laughren M, Hoogenraad R, Hale K & Japanangka Granites R (1996). *A learner's guide to Warlpiri*. Alice Springs: IAD Press.
McConvell P & Laughren M (2004). 'The Ngumpin-Yapa Subgroup.' In Bower C & Koch H (eds.) *Classification and the comparative method*. Amsterdam: John Benjamins. 169–196.
Nash D (1982). 'Warlpiri verb roots and preverbs.' In Swartz (ed.). 165–216.
Nash D (1986). *Topics in Warlpiri grammar*. New York/London: Garland Publishing.
Simpson J (1991). *Warlpiri morpho-syntax: a lexicalist approach*. Dordrecht/Boston/London: Kluwer Academic.
Simpson J, Nash D, Laughren M, Austin P & Alpher B (eds.) (2001). *Forty years on: Ken Hale and Australian languages*. Canberra: Pacific Linguistics.
Swartz S M (ed.) (1982a). *Work Papers of SIL-AAB, Series A. Vol. 6: Papers in Warlpiri grammar: in memory of Lothar Jagst*. Darwin.
Swartz S M (1982b). 'Syntactic structure of Warlpiri clauses.' In Swartz (ed.). Darwin: Summer Institute for Linguistics. 69–127.
Swartz S M (1991). *Constraints on zero anaphora and word order in Warlpiri narrative texts*. Darwin. SIL-AATB, Occasional Papers 1. Darwin: Summer Institute for Linguistics.

Welsh

P W Thomas, School of Welsh, Cardiff University, Wales, UK

© 2006 Elsevier Ltd. All rights reserved.

Demographic Features

According to the 2001 census, 582 000 people, or 20.8% of the total population of Wales, claimed to be able to speak Welsh; almost 798 000, or 28.4% of the population, claimed to have at least one language skill in Welsh. Before 2001 successive censuses had recorded a decline in the number of Welsh speakers: 10 years previously, in 1991, there were 508 000 speakers who comprised 18.7% of the population. Apart from some very young children, all Welsh speakers in Wales are bilingual and can also speak English.

There are no official counts of Welsh speakers outside Wales but surveys commissioned by S4C, the Welsh television channel, suggest that there are more than 200 000 in England. One particular area where emigrants from Wales continue to speak Welsh is Patagonia, in the Chubut province of Argentina. The first settlers arrived in 1865, hoping to found a 'New Wales'; many of their descendants are bilingual in Welsh and Spanish.

The density and numbers of Welsh speakers show considerable geographic variation. For example, 69% of the population of Gwynedd, in the north-west, can speak the language, compared to 11% of the population of Cardiff, the capital, in the south-east. But while the 69% of Gwynedd represents almost 78 000 speakers, the 50% of Carmarthenshire in the south-west represents more than 84 000 individuals. There are also more Welsh speakers in urban than rural areas. For example, there are almost 26 000 speakers in rural Powys in mid Wales, but almost 28 000 in the post-industrial Rhondda-Cynon-Taf valleys, 29 000 in the southern city of Swansea, and 32 500 in Cardiff, the capital.

The increase in Welsh speaking is the result of growth on two main fronts. The first, and most obvious, is among school children. In addition to the growing number of Welsh-medium schools, it became compulsory in 1990 for children in English-medium state schools to learn Welsh up to the age of 14; in 1999 the upper age limit was raised to 16. These changes are reflected in the 2001 census, which recorded that 40.8% of all children between the ages of 5 and 15 could speak Welsh. The second growth area in the number of Welsh speakers is the many thousands of adults who are learning the language.

The change in the ability to use Welsh is accompanied by increasing institutional support. A Welsh television channel, S4C, was established in 1982, followed in 1998 by S4C Digital, which broadcasts over 80 hours of Welsh television a week. There are several local radio stations and a national Welsh language radio station, Radio Cymru, which broadcasts about 126 hours a week. Several hundred Welsh language books and periodicals are published a year and a network of some 50 local Welsh papers which are produced several times a year by volunteers.

The use of the Welsh language is promoted by the Welsh Language Board, a government-funded body that was established by the 1993 Welsh Language Act, which states that Welsh and English should be treated equally in the administration of justice and in public business. Public bodies in Wales must submit schemes to the Board that describe the provision they make for the language. The aims of the Board seem to have general support: according to a recent opinion poll 67% of the people in Wales thought that more should be done to promote the language.

Periods of Welsh

The periods of the development of Welsh are conventionally divided into Early Welsh (up to the end of the 8th century and represented by a few names), Old Welsh (from the 9th to the 11th centuries, represented by glosses and fragments of prose and verse), Middle Welsh (from the 12th to the 14th centuries and represented by a substantial body of prose and verse), and Modern Welsh.

Linguistic Features

Alphabet

Written Welsh uses the Roman alphabet. Particular orthographic conventions include several digraphs, e.g. <th> for /θ/, <dd> for /ð/, <ch> for /χ/ and <ll> for /l/; <w> and <i> represent the consonants /w/ and /j/ or the vowels /u/ and /i/ respectively, and <y> represents /ə/ and /i(:)/.

Phonemic Inventory

The vowels, diphthongs, and consonants of Welsh are listed in **Table 1**. The main dialect variations with respect to this inventory are the absence of /i(:)/ (and diphthongs closing to /i/) in southern Welsh, of /ə/ in the extreme south-west, and of /h/ and voiceless /t/ in the south-east. Conservative northern speakers will substitute /s/ for /z/, which features in some loans

Table 1 Phoneme inventory of modern Welsh

| Vowels | | | | | | | | | |
|----------------------|----------|-------------|--------|----------|----------------|---------|-------|--------|---------|
| i | ĩ | u | i: | ĩ: | u: | | | | |
| e | ə | o | e: | o: | | | | | |
| | a | | | a | | | | | |
| Diphthongs | | | | | | | | | |
| iu | | eu | | | | | | | |
| ɛu | əi, əu | ɔĩ | | | | | | | |
| | ai, au | | | | | | | | |
| Consonants | | | | | | | | | |
| | Bilabial | Labiodental | Dental | Alveolar | Palatoalveolar | Palatal | Velar | Uvular | Glottal |
| Voiceless stops | p | | t | | | | k | | |
| Voiced stops | b | | d | | | | g | | |
| Voiceless fricatives | | f | θ | s, ʃ, ʁ | ʃ | | | χ | h |
| Voiced fricatives | | v | ð | z | (ʒ) | | | | |
| Nasals | m | | n | | | | ŋ | | |
| Liquids | | | l | r | | | | | |
| Semivowels | w | | | | | j | | | |

Table 2 Three dialect differences

| Feature | Example | Northern | Southern |
|---|-----------------|----------|----------|
| Realisation of /χ/ preceding /w/ | (ch)with 'left' | χwi:θ | (h)wi:θ |
| Vowel length preceding /t/ | call 'sane' | kaɫ | ka:tɫ |
| Vowel length preceding /sp, st, sk, ʃt/ | gw(a)llt | gwa:ɫt | gwaɫt |

from English. Two affricates – /tʃ/ and /dʒ/ – feature in loanwords and in dialects. Other salient dialect differences are listed in **Table 2**.

Consonant Mutation

In common with the other Celtic languages, some of the initial consonants of Welsh words vary according to their grammatical context, for example:

/ka:θ/ <cath>
'cat'
/və ŋha:θ/ <fy nghath>
'my cat'
/i ga:θ/ <ei gath>
'his cat'
/i χa:θ/ <ei chath>
'her cat'

Such consonantal changes are traditionally called mutations. They may be triggered by a preceding word, such as the personal pronouns in the above examples, or by grammatical context. For example, the object of a verb will mutate but not the subject:

Gwelodd ddyn.
saw-PAST man
'he/she saw a man'

Table 3 The consonant mutations

| Radical | Soft | Nasal | Spirant |
|---------|------|-------|---------|
| /p/ | /b/ | /mh/ | /f/ |
| /t/ | /d/ | /nh/ | /θ/ |
| /k/ | /g/ | /ŋh/ | /χ/ |
| /b/ | /v/ | /m/ | |
| /d/ | /ð/ | /n/ | |
| /g/ | – | /ŋ/ | |
| /m/ | /v/ | | |
| /t/ | /l/ | | |
| /r/ | /r/ | | |

Gwelodd dyn.
saw-PAST man
'a man saw'

There are three mutations, which may affect up to nine consonants (**Table 3**).

Vocabulary

The core vocabulary of Welsh is Celtic, for example, *drws* 'door', *dyn* 'man', and *haul* 'sun'. There are some 800 loanwords from Latin, mostly borrowed during the Roman occupation (43–410 A.D.); many of these refer to architectural and religious innovations, for example, *eglwys* 'church' from Latin *ecclesia*, *ffenestr* 'window' from *fenestra*, and *pont* 'bridge' from *pontem*. There are also many thousands of loans from English. A very few of these may be dated to the Old English period, but the numbers increase from the medieval period onward; examples are *cwpan* 'cup', *sêd* 'seat', *trowsus* 'trousers'.

There are some dialect differences in the vocabulary, particularly between northern and southern

varieties; for example, 'grandmother' is *nain* in northern Welsh but *mam-gu* in southern Welsh; 'out' is *allan* in the north but *mâs* in the south; and 'with' is *efo* in the north but *gyda* in the south. Standard Welsh may use both *nain* and *mam-gu*, but only *allan* and *gyda*. Speakers are generally tolerant of such variation.

Keeping pace with developments in English vocabulary has occupied lexicographers since the 18th century. More recently, educationalists who are concerned with delivering the school curriculum through the medium of Welsh have planned the elaboration of Welsh vocabulary through coinage, borrowing, and adaptation. The standardization of subject-specific vocabularies is undertaken professionally.

Syntax

Welsh is a VSO language. For example:

Prynodd y ferch gar.
bought the girl car.
'the girl bought a car.'

Welsh has a definite article but no indefinite article. Adjectives tend to follow the noun they qualify, for example:

car coch
car red
'red car'

Welsh has grammatical gender. Some adjectives have feminine and plural forms, a feature that is more prominent in formal styles and northern dialects, for example:

ceffyl gwyn
horse white
'white horse'

caseg wen
mare white-FEMININE
'white mare'

ceffylau gwynion
horses white-PLURAL
'white horses'

Numerals have masculine/neutral and feminine forms for 1 (*un*), 2 (*dau*, *dwy*), 3 (*tri*, *tair*) and 4 (*pedwar*, *pedair*). The gender of the numeral *un* is apparent only when nouns beginning with certain consonants follow it; cf.

ci, un ci
dog-MASCULINE, one dog
cath, un gath
cat-FEMININE, one cat

Stylistic Variation

Informal spoken varieties of Welsh show considerable variation, and may be heavily influenced by English vocabulary, morphology, syntax, and intonation, with frequent code-switching. Formal varieties tend to be more conservative and to favor native features.

Bibliography

- Aitchison J W & Carter H (1994). *A geography of the Welsh language 1961–1991*. Cardiff: University of Wales Press.
- Aitchison J & Carter H (2000). *Language, economy and society: the changing fortunes of the Welsh language in the twentieth century*. Cardiff: University of Wales Press.
- Aitchison J & Carter H (2004). *Spreading the word: the Welsh language 2001*. Talybont: Y Lolfa.
- Awbery G M (1984). 'Welsh.' In Trudgill P (ed.) *Language in the British Isles*. Cambridge: Cambridge University Press. 259–277.
- Ball M J & Jones G E (eds.) (1984). *Welsh phonology*. Cardiff: University of Wales Press.
- Davies J (1999). *The Welsh language*. Cardiff: University of Wales Press & The Western Mail.
- Griffiths B (ed.) (1995). *The Welsh Academy English–Welsh dictionary*. Cardiff: University of Wales Press.
- Jones M & Thomas A R (1977). *The Welsh language: studies in its syntax and semantics*. Cardiff: University of Wales Press.
- Parry-Williams T H (1923). *The English element in Welsh*. London: The Honourable Society of Cymmrodorion.
- Price G (1984). 'Welsh.' In *The languages of Britain*. London: Arnold. 94–133.
- Prys D & Jones J P M (eds.) (1998). *Y termiadur ysgol, Standardized terminology for the schools of Wales*. Bangor: University of Wales.
- Thomas A R (1973). *The linguistic geography of Wales*. Cardiff: University of Wales Press.
- Thomas A R (ed.) (2000). *The Welsh dialect survey*. Cardiff: University of Wales Press.
- Thomas R J, Bevan G A & Donovan P J (eds.) (1950–2002). *Geiriadur Prifysgol Cymru, A dictionary of the Welsh language*. Cardiff: University of Wales Press.
- Watkins T A (1993). 'Welsh.' In Ball M J & Fife J (eds.) *The Celtic languages*. London: Routledge. 289–348.

Relevant Website

<http://www.bwrdd-yr-iaith.org.uk/> – Welsh Language Board.

West Greenlandic

A Berge, University of Alaska, Fairbanks, AL, USA

© 2006 Elsevier Ltd. All rights reserved.

The Language and Its Dialects

Greenlandic, or Kalaallisut, is an Eskimo language (*see Eskimo-Aleut*). Greenland, or Kalaallit Nunaat, is geographically and culturally part of the North American continent; however, since 1721, it has been a territory of Denmark. In 1979 Greenland obtained autonomy over local governance, and Greenlandic was named the national language along with Danish. Today there are more than 50 000 speakers of Greenlandic, the vast majority of whom live in Greenland, although a sizable population is to be found in Denmark. There are three major dialects of Greenlandic: Polar Eskimo, spoken in the Thule region, East Greenlandic, spoken on the east coast, and West Greenlandic, spoken along most of the western coast. West Greenlandic is the dialect most widely spoken in Greenland, as well as being the standard dialect for purposes of political administration, education, church, and media. The dialect region stretches from Upernavik in the north to Kap Farvel in the south. Four subdialects are generally recognized: the subdialect spoken in Upernavik; North West Greenlandic (Uummannaq, and the Disko Bay region), Central West Greenlandic (Sisimiut to south of Nuuk); and South West Greenlandic (from north of Paamiut to south of Nanortalik), according to Dorais, 1996. The subdialects differ slightly in various aspects of their phonology and lexicon; thus, the Upernavik dialect, sharing a feature of East Greenlandic, tends to replace /u/ with /i/ under certain conditions, and Northwest Greenlandic retains a historical retroflex /ɰ/, whereas the Central and Southwestern dialects have merged /ɰ/ with /s/. The Greenlandic spoken in the capital city, Nuuk, tends to contain more Danish loans and syntactic features than the language spoken in other settlements, due to a relatively significant Danish population, as well as to its concentration of administrative and political activities. The Central West Greenlandic subdialect has long been the accepted spoken and written standard, and it will serve as the basis of the description given below.

Historiography of Descriptive Work

Greenland has seen several waves of immigration from both Eskimo and European populations. The most recent Eskimo groups are estimated to have

arrived in Northern Greenland by the end of the 12th century and in Southern Greenland by the end of the 15th century. There is some evidence they made contact with the first European immigrants, the Norse, who had arrived in the late 10th century. There is, however, scant linguistic evidence of Norse influence on Greenlandic, and the Norse had disappeared by the 16th century. From the late Middle Ages, European whalers, traders, and explorers made their way up the coast of Greenland; the first written records of West Greenlandic are wordlists they compiled during the 16th and 17th centuries, although some of the words collected appear to represent a trade pidgin (van der Voort, 1996).

The first systematic grammatical descriptions of the language date from the beginning of the Danish colonial era in the 18th century and were made by Lutheran and Moravian missionaries. The first of these was a collaborative work from 1725 between the missionary Hans Egede and his assistant Topp, with the help of Egede's son, Poul. This served as the basis of a dictionary, published in 1750, and the first complete grammar, in 1760, by P. Egede. Later descriptions were modeled on P. Egede's work, the most notable being O. Fabricius' grammar (1791, rev. 1801) and dictionary (1804). In 1851 Samuel Kleinschmidt published his grammar of Greenlandic; this is the earliest thorough linguistic description of an American native language and it is widely seen as the first modern, synchronic linguistic description of a language. Kleinschmidt also created a standard, linguistically accurate orthography for Greenlandic, which was maintained until 1972, when a more modern orthography was introduced to reflect important morphophonological changes in Greenlandic. Descriptive work on West Greenlandic has continued to the present, and important scholars include Thalbitzer, one of the first to document through sound recordings, Schultz-Lorentzen, Bergsland, and Fortescue. In addition to general descriptions, more recent work has included specialized studies of phonology (e.g., Rischel, 1972), syntax (e.g., Sadock, 1991), and discourse (e.g., Berge, 1997).

Phonetics and Phonology

West Greenlandic phonology is characterized by having few consonants and vowels and restrictions on vowel and consonant clusters, as well as on final consonants. Thus, only vowels and stops are found word finally; the only allowable diphthongs are /iV/, /Vi/, or /uV/; and consonant clusters are only found medially

Table 1 Phoneme inventory for West Greenlandic

| Manner/Place | Labial | Dental | Palatal | Velar | Uvular |
|--------------------------------|--------|--------|-----------|-------|--------|
| Stops | p | t | | k | q |
| Fricatives –v | f | s | (ʃ = [s]) | x | |
| + v | v | | | g | R |
| Nasals | m | n | | ŋ | |
| Liquids | | | ɬ = [ll] | | |
| | | | l | | |
| Glides | j | | | (w) | |
| Vowels: a, aa, i, ii, u, uu | | | | | |

Standard orthography is in brackets; parentheses indicate subdialect forms or questionable phonemic status.

and consist mostly of geminates, with the exception of /rC/ combinations (see Table 1).

Some features of Greenlandic are common to other Eskimo languages, including traces of a fourth vowel (see *Eskimo-Aleut*), and a rich morphophonology. Particular to West Greenlandic is the extreme degree of consonant cluster assimilation, which has taken place in the historic period.

| | |
|-----------------|--------------------------------------|
| Old Greenlandic | paurqi-ngnig-tar-fik |
| New | paaqqi-nnit-tar-fik |
| | <i>take.care.of-ANTI-HABIT-place</i> |
| | <i>'nursing home'</i> |

Restrictions on vowel or consonant clusters and a productive morphology, with complex rules for adding morphemes, have led to more opaque word formations than in other Eskimo languages. (For more on the morphophonology, see Rischel, 1974 and Fortescue, 1984.)

Morphology/Syntax

Greenlandic is an extremely polysynthetic language, with a large number of derivational affixes and complex inflection. Words consist of a root (or base), typically from zero to five suffixes known as postbases (although more than five are possible and quite normal), and an inflectional ending; with one non-productive exception, there are no prefixes. Roots generally are subcategorized for part of speech; nominal roots will require nominal inflection, and verbal roots will require verbal inflection. The most important parts of speech are the open classes of nouns and verbs; adjectives and adverbs tend to be verbally derived. There are also a rich system of demonstratives, a limited set of particles, a limited set of fossilized adverbs and adjectives, and few but common clitics. There are several hundred derivational postbases, many of which are highly productive. These are commonly classified into four categories: those

which derive nouns from nominal bases (NN); those that derive verbs from nominal bases (NV); those that derive verbs from verbal bases (VV); and those that derive nouns from verbal bases (VN). Some also attach to other parts of speech, e.g., particles; *nnit*- and *tar*-in the example above are both VV, and *-fik* is VN. NN and NV are shown in the following example:

inuk-rsuaq-u-voq
man-big-COP-3sing.INDIC
 N-NN-NV-INFL
 'he is a giant'

As this example shows, verbalizing postbases can create verbal structures that 'incorporate' a verbal argument; that is, a subject or object can be brought into the verbal structure. There is some theoretical debate as to whether or not Eskimo languages can be called incorporating (e.g., Baker, 1988), but within the field of Eskimo linguistics, there is a long tradition of using the term 'incorporation' for structures of the type exemplified above. Even inflected forms can incorporate:

aappalut-toq illu-mi-iC-voq
red-PART house-LOC-COP-3sing.INDIC
 'he is in the red house'

Nominal inflection includes eight cases as well as possessive and person markers, which are also inflected for case. There are two grammatical cases, absolutive and ergative (or relative), and six oblique cases, including instrumental (a default case with many functions, both grammatical and nongrammatical), locative, ablative, allative, vialis, and equalis.

Verbal inflection is semifused and includes marking for dependence, mood, transitivity, person, and number. Verb moods are typically categorized as either independent or dependent. Sentences often consist of strings of subordinate clauses with dependent mood marking and a superordinate clause headed by a verb with independent mood marking. Independent moods include the indicative, interrogative, optative, and imperative moods. Dependent moods include the conditional, causative (indicating causation or action prior to that expressed by the independent clause), contemporative (indicating action contemporaneous with that expressed by the independent clause), and participial (used in object clauses, as an alternative to the indicative in narration, and in other discourse contexts). For an example of clause chaining from West Greenlandic (see *Eskimo-Aleut*).

Ergativity and transitivity have long been topics of interest in the study of Greenlandic. Transitive



Figure 1 Major Greenlandic dialects and West Greenlandic subdialects.

clauses are headed by verbs with subject and object person marking, and the arguments are marked with ergative and absolutive case. Intransitive and antipassive clauses are headed by verbs with pronominal marking of the subject and nouns take absolutive case. Objects of antipassive clauses take instrumental case. Traditionally, these are seen as related to definiteness of the object, although they may reflect topicality (Berge, 1997).

| | | |
|--|----------------------|----------------------------|
| anguti-p | nanuq | taku-aa |
| <i>man-ERG.sing</i> | <i>bear.ABS.sing</i> | <i>see-3SG.3sing.INDIC</i> |
| 'the man sees/saw the bear' (definite, or old topic) | | |

| | | |
|---|-----------------------|------------------------|
| angut | nanur-mik | taku-voq |
| <i>man.ABS.sing</i> | <i>bear-INST.sing</i> | <i>see-3sing.INDIC</i> |
| 'the man sees/saw a bear' (indefinite, or topic introduction) | | |

At least one relatively major syntactic theory has been developed to account for Greenlandic morphosyntax.

This is Autolexical Theory, developed by Sadock (1991, 2003) and based in part on GPSG. It allows morphology and syntax to require structures that may not always result in exact matching, as in the stranded modification of a locative phrase in the sentence 'he is in the red house' given above.

Lexicon

Modern Greenlandic has seen an influx of new lexical items as a result of colonial experience and modernization. The earliest evidence of this is seen in early Bible translations, with the heavy use of Danish loans relating to Christianity. Some of these loans have been well integrated into the language, for example, *palasi*, from Danish *præst* 'priest', while others have been replaced by Greenlandic coinages. Greenlandic has actively incorporated new words in its lexicon, through relexicalization of obsolete terms (e.g., *issat* 'snow goggles' are now 'eyeglasses'), borrowing (*kaffi* 'coffee'), and coinage (Petersen, 1976, Berge and Kaplan, 2005). There is some evidence for the increased use of passive formations and nominalizations in the lexicon, perhaps as a result of the influence of journalistic style and literacy.

Semantics/Discourse/Sociolinguistics

Little work has been done to date on other aspects of linguistic description, especially including semantics, discourse, and sociolinguistics, although the body of work in these areas is steadily increasing. There have been reports on child language acquisition (Fortescue, 1985), bilingualism (Jacobsen, 1997), and discourse (Berge, 1997), among others. More sociolinguistic and semantic studies have been done of neighboring dialects, such as Inuktitut.

State of the Language Today

Unlike many other native languages of North America, West Greenlandic is not endangered and is, in fact, undergoing normal language development and change, although there was a period of endangerment. During the mid-20th century, increasing encroachment of the Danish language led to almost two generations of speakers who appeared to be losing their fluency in Greenlandic. In response, Greenlandic became one of the national symbols of the political campaigns for autonomy from Danish rule in the 1970s. With the establishment of Home Rule in 1979, Greenlandic was given the status of national language along with Danish. Language loss was successfully reversed, although there may have

been some lasting effects of bilingualism on the dialect spoken in Nuuk. Factors that have contributed to this reversal include a history of education and literacy in Greenlandic, a wealth of materials in the language, and political support. From earliest colonial times, education was established in West Greenlandic, and literacy was common. Today Greenlandic can be chosen as a medium of instruction throughout the years of formal schooling, and there have been efforts to include it as the language of instruction at Ilisimatusarfik, the University of Greenland. The first books in Greenlandic were published in the 1850s and the first newspaper, *Atuagagdliutit*, shortly thereafter. Since then, there have been several thousand books and articles published in Greenlandic, *Atuagagdliutit* has been in continuous print since its inception, and there are today two bilingual newspapers (in Danish and Greenlandic), local television and radio stations with Greenlandic programming, and more.

Bibliography

- Baker M (1988). *Incorporation: a theory of grammatical function changing*. Chicago: University of Chicago Press.
- Berge A (1997). 'Topic and discourse structure in West Greenlandic agreement constructions.' Ph.D. diss., University of California, Berkeley.
- Berge A & Kaplan L (2005). 'Contact-induced lexical development in Yupik and Inuit languages.' *Études/Inuit/Studies*.
- Bergsland K (1955). 'A grammatical outline of the Eskimo language of West Greenland.' Oslo: Mimeo.
- Dorais L-J (1996). *La parole Inuit: langue, culture, et société dans l'Arctique nord-américain*. Paris: Peeters.
- Egede P (1760). *Grammatica Grönländica Danico-Latina*. [Copenhagen]: Gottmann.
- Fabricius O (1801). *Forsøg af en forbedret Grønlandsk grammatica*. Copenhagen: Kongelige Waysenhuses Bogtrykkerie.
- Fortescue M (1984). *West Greenlandic*. London: Croom Helm.
- Fortescue M (1985). 'Learning to speak Greenlandic: a case study of a two-year-old's morphology in a polysynthetic language.' *FL* 5, 101–114.
- Jacobsen B (1987). 'A preliminary report on a pilot investigation of Greenlandic school children's spelling errors.' In Luelsdorff P A (ed.) *Orthography and Phonology*. Amsterdam: John Benjamins. 101–130.
- Kleinschmidt S (1851). *Grammatik der Grönländischen Sprache*. Berlin: G. Reimer.
- Petersen R (1976). 'Nogle træk i udviklingen af det grønlandske sprog.' *Tidskriftet Grønland* 6, 165–208.
- Rischel J (1974). *Topics in West Greenlandic phonology*. Copenhagen: Akademisk Forlag.
- Sadock J M (1991). *Autolexical syntax: a theory of parallel grammatical representations for series* *Studies in*

- Contemporary Linguistics*. Chicago: University of Chicago Press.
- Sadock J M (2003). *A grammar of Kalaallisut. (West Greenlandic Inuttut)*. Muenchen: LINCOM.
- Schultz-Lorentzen C W (1945). *A grammar of the West Greenlandic language, for series Meddelelser om Grønland 129,3*. Copenhagen: C.A. Reitzels Forlag.
- Tersis N & Therrien M (eds.) (2000). *Les langues Eskaléoutes: Sibérie, Alaska, Canada, Groënland*. Paris: CNRS.
- Thalbitzer W (1911). 'Eskimo.' In *Handbook of Native American languages*. Bureau of Indian Ethnology Bulletin no. 40. Washington: Government Printing Office. 967–1069.
- van der Voort H (1996). 'Eskimo pidgin in West Greenland.' In Jahr E & Broch I (eds.) *Language contact in the Arctic: northern pidgins and contact languages*. Berlin: Mouton de Gruyter. 157–260.

West Papuan Languages

G Reesink, Leiden University, Leiden,
The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Introduction

In the area between Timor and the adjacent islands Alor and Pantar (126°E) and the Cenderawasih Bay of the Indonesian province Papua (136°E), roughly 50 of the more than 800 Papuan languages are spoken (see Figure 1).

These West Papuan languages do not form one family, in spite of earlier attempts to establish a "large West Papuan Phylum" (Cowan, 1957, 1960). More recently, the West Papuan Phylum has been restricted to the languages of North Halmahera (NH) and the Bird's Head Peninsula (Voorhoeve, 1975; Wurm, 1981, 1982), whereas the South Bird's Head (SBH) languages and some languages on the western tip of the Bomberai peninsula are claimed to form one family with those of Timor-Alor-Pantar (TAP), forming a subgroup within the largest Papuan family proposed thus far, the Trans New Guinea (TNG) family (McElhanon and Voorhoeve, 1970; Voorhoeve, 1975; Stokhof, 1975; Wurm, 1982; Pawley, 1998; Foley, 2000; Ross, 2004).

The West Papuan languages of North Halmahera and the Bird's Head could form a distantly related family, on the basis of (i) pronominal forms for 1sg as **t/d-* and **n-* for 2sg; (ii) the number-ablaut (sg is *a*, pl is *i*), found in TNG, is also attested in a number of West Papuan languages, and (iii) a small number of possible cognates. Some of these show some overlap with the TNG evidence, for example, reflexes of **niman* 'louse' (Pawley, 1998; Reesink, 2004). Although a few families have been established with reasonable certainty, the evidence for linguistic relatedness between them is so meagre that no firm conclusions should be drawn at present. Rather, the West Papuan languages form an areal network of basically

unrelated families (North Halmahera, West BH, and two East BH families (Meyah-Sougb; and Hatam-Mansim) and a number of isolates in the center of the Bird's Head (Maybrat, Abun, Mpur) and Yawa in the Cenderawasih Bay. They share a number of typological features, not only between them but also with the Austronesian languages spoken in this same region, betraying approximately four millenia of contact since the Austronesians first arrived in the Moluccas and around the Bird's Head (Bellwood, 1985).

Typological features

Verbal Complex

As typical of Papuan languages, the constituent order of the clause is SOV in TAP, NH, SBH, and Yawa, all of which also have a verbal prefix for (animate) object, which is the normal crossreference for the recipient with a verb like 'give.' Less typical is the configuration that has an additional verbal prefix for subject, as found in NH and SBH languages.

The region of the west Papuan languages is one of three in which Papuan languages are found that do not have a V-final order, the other two being the Torricelli languages and some of the East Papuan languages. Some of the NH and most BH languages have a rather strict SVO order, often with a subject prefix as the only verbal affixation.

Tense-Aspect-Mood morphology is generally poor or completely absent, as in Abun. Some aspect or mood prefixation is found in languages of the EBH. Inanwatan has tense marked by suffixation, whereas the TAP languages mark aspect that way. In the 'non-tensed' languages, predicative adjectives behave as verbs (see Stassen, 2003), whereas 'tensed' Inanwatan verbalizes adjectives by means of a copula (De Vries, to appear).

All West Papuan languages, whether OV or VO, have a clause-final negative adverb, in some cases with no morphosyntactic means to delineate its

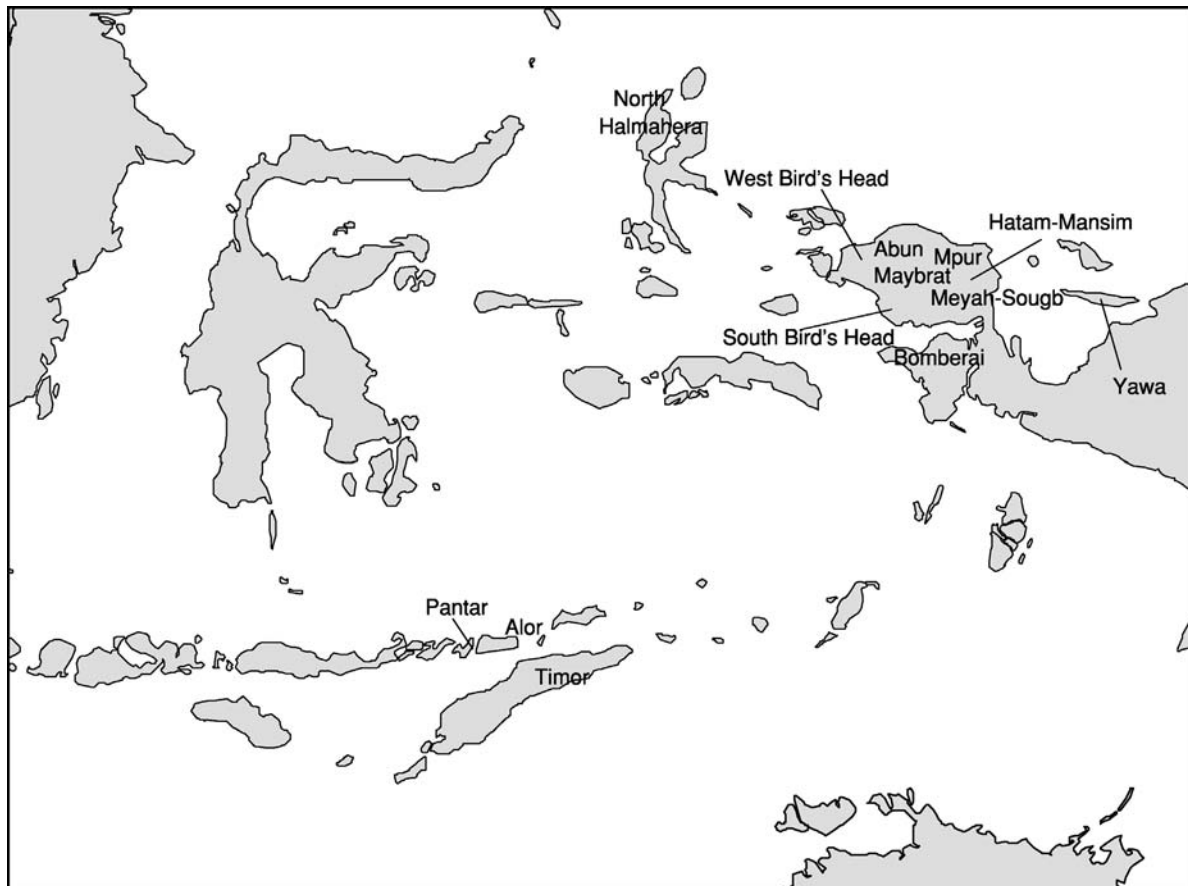


Figure 1 Map of West Papua.

scope (Reesink, 2002). They also agree in having clause-final aspectual adverbs, such as ‘already.’

Nominal Complex

The order of constituents in the Noun Phrase is in all West Papuan languages: Noun–Adjective–Numeral–Demonstrative, whereas a few NH languages have a prenominal article in addition. All of them make a distinction between alienable and inalienable possession. The latter construction consists of a possessor prefix on the possessed noun (body-part and kinship terms), which is generally identical to either the subject or object prefix, if the latter is available in the language.

Nominal classifiers are widely available in the West Papuan languages, Meyah having the most complex system, whereas its relative Sougb and their (very distantly?) related neighbour Hatam only have a vestige (see Reesink, 2002).

Pronominal systems A gender distinction (masculine; feminine; and, in some languages, neuter) for 3sg forms seems to be an old Papuan feature that links the West Papuan languages with most of the non-TNG languages along the north coast of New

Guinea, as far as the Solomon Islands. The TAP languages, and on the Bird’s Head, the isolate Abun and the two EBH families lack this feature.

The inclusive-exclusive opposition for nonsingular first person seems to be an Austronesian feature that has found its way into all the WP languages, except three central BH isolates, Maybrat, Abun, and Mpur.

Tone

Tonal contrasts are found in Mpur, with four phonemic tones (Odé, 2002) and Abun with three (Berry and Berry, 1999), whereas Meyah (Gravelle, 2002) and Sougb (Reesink, 2002) are pitch-accent languages with two contrastive tones. Ma’ya and Matbat are AN languages of the Raja Ampat Islands that have a Papuan substrate of four contrastive tones (Remijsen, 2001).

Papuan and Austronesian Contact

The SVO order with concomitant prepositions can be seen as a diffusion into many WP languages, in addition to the inclusive–exclusive opposition. There

seems to have been more diffusion in the other direction: almost all Austronesian languages in the WP sphere have a clause-final negator, a preposed possessor, and the alienable–inalienable distinction for possessive construction, albeit that the latter is expressed by possessor suffixes on the possessed noun, rather than by prefixes as in the Papuan languages. The typological contrast between western AN languages and those of this area, given by Himmelmann (to appear), focuses precisely on these ‘Papuanisms’ (see Klammer, Reesink, and Van Staden, to appear). It thus suggests a scenario of original Papuan-speaking communities that shifted to ‘imperfectly’ learned Austronesian languages. Prolonged contact between these communities allowed for further convergence to a linguistic area of AN and Papuan languages in the Moluccas and the western peninsula of New Guinea.

Bibliography

- Bellwood P (1985). *Prehistory of the Indo-Malaysian archipelago*. New York: Academic Press.
- Berry K & Berry C (1999). *A description of Abun, a West Papuan language of Irian Jaya*. Canberra: Pacific Linguistics.
- Cowan H J K (1957). ‘A large Papuan language phylum in West New Guinea.’ *Oceania* 28(2), 159–166.
- Cowan H J K (1960). ‘Nadere gegevens betreffende de verbreiding der West-Papoease Taalgroep.’ *Bijdragen Taal-, Land- en Volkenkunde* 116, 350–364.
- De Vries L (1996). ‘Notes on the morphology of the Inanwatan language.’ *NUSA* 40, 97–127.
- Foley W A (2000). ‘The languages of New Guinea.’ *Annual Review of Anthropology* 29, 357–404.
- Gravelle G (2002). ‘Morphosyntactic properties of Meyah word classes.’ In Reesink (ed.) 109–180.
- Himmelmann N (in press). ‘The Austronesian languages of Asia and Madagascar: typological characteristics.’ In Adelaar K A & Himmelmann N P (eds.) *The Austronesian languages of Asia and Madagascar*. London: Routledge/Curzon.
- Klammer M, Reesink G P & Van Staden M (to appear). ‘East Nusantara and The Bird’s Head as a linguistic area.’ In Muysken P (ed.). *Volume on Linguistic areas*.
- McElhanon K A & Voorhoeve C L (1970). *The Trans-New Guinea Phylum: explorations in deep-level genetic relationships*. Canberra: Pacific Linguistics.
- Miedema J, Odé C & Dam R A C (eds.) (1998). *Perspectives on the Bird’s Head*. Amsterdam: Rodopi.
- Odé C (2003). *Mpur prosody: an experimental–phonetic analysis with examples from two versions of the Fentora myth*. Osaka: Endangered Languages of the Pacific Rim [ELPR Publication series A1–003].
- Pawley A K (1998). ‘The Trans New Guinea Phylum hypothesis: a reassessment.’ In Miedema, Odé & Dam (eds.). 655–690.
- Pawley A K, Attenborough R, Golson J & Hide R (eds.) *Papuan Pasts: studies in the cultural, linguistic and biological history of the Papuan-speaking peoples*. Adelaide: Crawford House Australia.
- Reesink G P (1996). ‘Morpho-syntactic features of the Bird’s Head languages.’ *NUSA* 40, 1–20.
- Reesink G P (1998). ‘The Bird’s Head as Sprachbund.’ In Miedema, Odé & Dam (eds.). 603–642.
- Reesink G P (ed.) (2002). *Languages of the eastern Bird’s Head*. Canberra: Pacific Linguistics.
- Reesink G P (2002). ‘Clause-final negation: structure and interpretation.’ *Functions of Language* 9(2), 239–268.
- Reesink G P (2004). ‘Roots and development of West Papuan languages: Papuan elements in West Papuan languages.’ In Pawley, Attenborough, Golson & Hide (eds.).
- Remijsen B (2001). *Word-prosodic systems of Raja Ampat languages*. Utrecht: LOT (Netherlands Graduate School of Linguistics).
- Ross M (2004). ‘Pronouns as markers of genetic stocks in non-Austronesian languages of New Guinea and Island Melanesia and Eastern Indonesia.’ In Pawley, Attenborough, Golson & Hide (eds.).
- Stassen L (1997). *Intransitive Predication*. Oxford: Oxford University Press.
- Stokhof W A L (1975). *Preliminary notes on the Alor and Pantar languages (East Indonesia)*. Canberra: Pacific Linguistics.
- Voorhoeve C L (1975). *Languages of Irian Jaya: checklist, preliminary classification, language maps, wordlists*. Canberra: Pacific Linguistics.
- Wurm S A (ed.) (1975). *New Guinea area languages and language study, vol. 1: Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics.
- Wurm S A (1981). ‘Papuan language stocks, western New Guinea area.’ In Wurm S A & Hattori S (eds.) *Language atlas of the Pacific area, part 1: New Guinea area, Oceania, Australia*. Canberra: Pacific Linguistics.
- Wurm S A (1982). *The Papuan languages of Oceania*. Tübingen: Gunter Narr.

Wolaitta

A Amha, Leiden University, Leiden, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The term ‘Wolaitta’ (Wolaytta) designates both the speakers and the language discussed in this article. Their administrative unit, known as the Wolaitta Zone, is part of the Southern Peoples, Nations and Nationalities Regional State of Ethiopia. The northern neighbors of the Wolaitta are the Kambatta (Kambaata) and Hadiyya (Cushitic); the southern neighbors are the Gamo and Gofa peoples (Omotic). The western and eastern parts of Wolaitta are bounded by the Omo and Bilate rivers, respectively. Most of the Wolaitta are farmers. According to the 1994 national census of Ethiopia, there are 1 210 000 Wolaitta speakers. For the names of closely related languages, see the language family tree in **Figure 1**.

The Sound System

Consonants

Table 1 lists a consonant inventory of Wolaitta. [p] and [ɸ] are free variants in word-initial and intervocalic positions; [ɸ] does not occur as a geminate or as a member of a consonant cluster. The labial implosive is attested both in word-initial and medial positions, as in *ḥánk’a* ‘very sour’ and *šobḥá* ‘armpit,’ and the alveolar implosive occurs only in word medial position, as in *šódḏe* ‘frog.’ [ʒ] is used marginally and only in ideophonic words. Gemination is contrastive.

Vowels

Wolaitta has a five-vowel system with each vowel having a longer counterpart (Table 2). Examples are *mára* ‘calf,’ *maára* ‘row’ and *boóra* ‘ox,’ *bóra* ‘critic.’

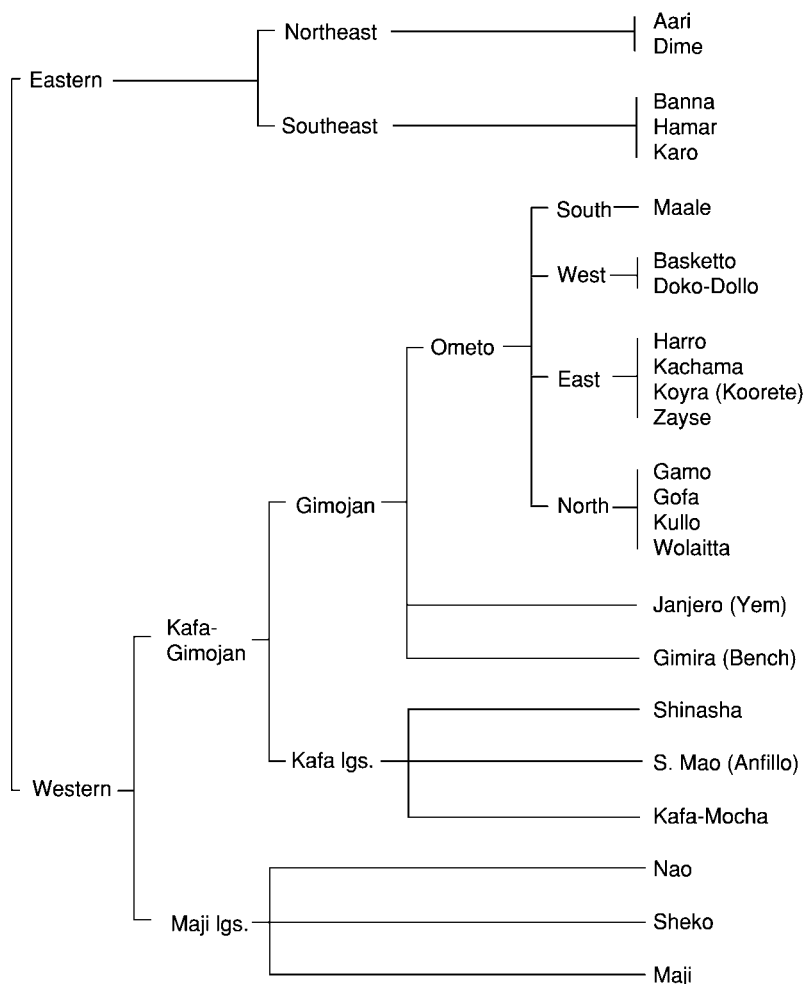


Figure 1 Omotic family tree, based on Fleming (1976).

Table 1 Consonant inventory

| | | | | |
|-----|----|-----|----|---|
| (p) | t | c | k | |
| b | d | j | g | |
| | t' | c' | k' | ʔ |
| ɸ | d' | | | |
| m | n | | | |
| (Φ) | s | š | | h |
| | z | (ž) | | |
| | l | | | |
| | r | | | |
| w | | y | | |

Table 2 Vowel inventory

| | |
|-------|-------|
| i, ii | u, uu |
| e, ee | o, oo |
| | a, aa |

Syllable Structure

CV, CVV, CVC, and CVVC syllable types are attested, as in *tá* 'I', *ʔéé* 'yes,' *mal.dó* 'sorghum,' and *keet.tá* 'house' (where the period indicates the syllable break). As the syllabification of the words *maldó* and *keettá* demonstrates, geminates and consonant clusters are split between two different syllables; also, clusters and geminates consist of only two members and they occur only word-medially. Syllable nucleus may be simple or branching.

Tone-Accent

Wolaitta is a tone-accent language. The language has two tones (high and low) used for making lexical distinction, as in *góda* 'lord, chief' versus *godá* 'wall' (where high tone is marked with ^ˈ and low tone is not marked). With a few exceptions (e.g., *ha* 'this,' *ta* 'my'), there are no words with just low tones; instead, lexical items have at least one high tone, mainly occurring on the ultimate or penultimate vowel. There are, however, a few numerals and nouns with ante-penultimate high tone, for example, *másunta* 'wound' and *k'éretta* 'split wood,' which seem to be historically derived from complex forms.

Nouns

Basic nouns in Wolaitta end in one of the following vowels: [e], [o], or [a] (Table 3). Which of these vowels a particular word may take cannot be predicted. There are no nouns ending in [i] or [u] in Wolaitta, although such nouns are attested in related languages.

Table 3 Basic nouns

| [e]-ending | [o]-ending | [a]-ending |
|-------------|------------------|----------------|
| búhe 'dust' | káwo 'dinner' | šaáfa 'river' |
| molé 'fish' | šooró 'neighbor' | keettá 'house' |

Table 4 Plural marking

| Singular | Nominative plural | Accusative plural | Gloss |
|----------|-------------------|-------------------|------------|
| šaáfa | šaáfa-t-i | šaáfa-t-a | 'river' |
| šooró | šooro-t-i | 'šooro-t-á' | 'neighbor' |

Plural Marking

On definite nouns, plural is marked by the morpheme *-t-*; indefinite nouns are not marked for plurality. Singular is unmarked. Examples are in Table 4.

Case, Gender, and Definiteness

Case, gender, and definiteness are designated cumulatively by portmanteau morphemes. In animate nouns, gender is determined by sex. Inanimate nouns are generally inflected like masculine nouns; but, when a diminutive meaning is intended, they may be inflected as feminine nouns. Plural nouns take the same nominative and accusative case markers as masculine singular nouns. Examples are in Table 5.

The genitive case is marked by *-ee* in definite feminine nouns and by *-u* in plural nouns. In masculine nouns, the genitive and accusative cases are formally identical. The possessor noun always precedes the possessed noun. Consider the forms of *šooró* 'neighbor' and *goššá* 'farm' in (1).

- (1a) *šooró gošša* 'a neighbor's farm'
 (1b) *šoor-úwa gošša* 'the neighbor's (MASC) farm'
 (1c) *šoor-eé gošša* 'the neighbor's (FEM) farm'
 (1d) *šooro-t-ú gošša* 'the neighbors' farm'

Peripheral/semantic cases such as instrumental (*-ra*), ablative (*-ppe*), dative (*yo/-ssi*), and so on are attached to a noun already marked with the genitive (for feminine and plural nouns) or accusative (for masculine singular nouns). Compare the examples in (1) with those in Table 6.

Nominal Derivation

There are several productive derivational suffixes, for example, *-ta* in *laggé-ta* 'friendship' (*lágge* 'friend') and *-tétta* in *zoʔó-tétta* 'redness' (*zoʔó* 'red'). Suffixing *-anqa* to a noun may derive agent

Table 5 Definiteness, case, and gender inflection

| | Basic noun | Definite masculine singular | | Definite feminine singular | | Definite plural | |
|----------------|--------------------------------|-----------------------------|------------------|----------------------------|--------------------|-----------------|------------|
| | | NOM | ACC | NOM | ACC | NOM | ACC |
| /a/- ending | keettá 'house' naʔá 'child' | keettáy naʔáy | keettáa naʔaa | keettiya naʔiya | keettiyo naʔiyo | keetta-t-i | keetta-t-a |
| /o/- ending | migído 'ring' | migídoy | migíduwa | migídiya | migídiyo | migído-t-i | migído-t-á |
| /e/- ending | šooró 'neighbor' | šooróy | šoorúwa | šooriya | šooriyo | šooro-t-i | šooro-t-á |
| | šóddé 'frog' | šóddée | šóddiya | šóddiya | šóddiyo | šóddé-t-i | šóddé-t-á |
| | zeére 'orphan' | zeéree | zeériya | zeériya | zeériyo | | |

Table 6 Peripheral semantic cases

| | 'from a' | | 'with a' | |
|----------|--|--|---------------------------------------|--|
| | Indefinite | Definite | Indefinite | Definite |
| Singular | šooró-ppe 'from a neighbor (MASC/FEM)' | šoor-úwa-ppe 'from the neighbor (MASC)' šoor-eé-ppe 'from the neighbor (FEM)' | šooró-ra 'with a neighbor (MASC/FEM)' | šoor-úwa-ra 'with the neighbor (MASC)' šoor-eé-ra 'with the neighbor (FEM)' |
| Plural | | šooro-t-ú-ppe 'from the neighbors' | | šooro-t-uú-ra 'with the neighbors' |

Table 7 Derivational suffixes

| Noun | Agent nominal | Adjective |
|--------------------|--------------------------|-----------------------------|
| kiita 'message' | kiit-ánca 'messenger' | |
| ʔóla'a 'fight/war' | ʔol-ánca 'fighter' | |
| dooná 'mouth' | | doon-aáma 'talkative' |
| wolk'á 'power' | | wolk'-aáma 'one with power' |

nouns; whereas suffixing *-aáma* to a noun derives an adjective (Table 7).

Adjectives and Adverbs

Adjectives end in one of the word-final vowels, *e*, *o*, or *a*. When used as modifiers, adjectives are not marked for gender, case, or number; they generally do not show agreement with the head noun. However, when the head noun is dropped, the adjective must be marked for these categories.

- (2a) keéha 'kind'
 (2b) keéha šooro-y 'the kind neighbor (NOM)'
 (2c) keéha-y 'the kind one'

In the inchoative, the adjectival base is affixed with tense-aspect and mood markers, as in:

- (3a) keeh-iísi 'he became kind'
 (3b) keeh-aásu 'she became kind'
 (3c) keeh-íbénna 'he did not become kind'

Manner adverbs are mainly derived by suffixing the locative marker *-n*, the instrumental *-ra*, or the ablative *-ppe* to nominals. For example:

- (4a) ʔakeéka-ni ʔoott-á
 attention-LOC do-2.SING.IMP
 'work carefully!'

- (4b) keeh-í-ppe harg-eési
 be_kind-í-ABL be_sick-3.MASC.SING.IMPERF
 'he is extremely/badly sick'

- (4c) ʔiss-í-ppe y-iite
 one-í-ABL come-2.PL.IMP
 'come together!'

Lexical time-adverbs include *baʔʔí* 'now,' *kasé* 'earlier,' *háčči* 'today,' and *wontó* 'tomorrow.'

Pronouns

The basic pronoun paradigms of Wolaitta are possessive, nominative, and accusative. Dative, ablative, and locative pronouns are formed by adding the respective case suffixes (i.e., *-ssi*, *-ppe* and *-n(i)*, as in the nouns) to the accusative/possessive ones (see Table 8). Note the gender syncretism between third-person singular pronouns in the forms in Table 8. The pronoun sets with *ba(-)* are used when the subject of the sentence is coreferential with an object or possessive noun in the same sentence, as shown in (5a), which contrasts with the noncoreferential form in (5b).

Table 8 Pronouns

| Person ^a | Possessive | Nominative | Accusative | Dative | Ablative |
|---------------------|------------|------------|------------|----------|----------|
| 1 SING | ta | tááni/tá | tána | taássí | taáppé |
| 2 SING | ne | nééni/né | néná | neéssí | neéppé |
| 3 FEM SING | ʔi | ʔá | ʔó | ʔíssí | ʔíppé |
| 3 MASC SING | ʔa | ʔi | ʔá | ʔássi | ʔáppé |
| 1 PL | nu | núúni/nú | núná | nuússi | nuúppé |
| 2 PL | ʔinte | ʔinté | ʔintena | ʔintéssi | ʔintéppé |
| 3 PL | ʔeta | ʔeti | ʔetá | ʔetássi | ʔetáppé |
| 3 SING LOG | ba | — | báná | baássi | baáppe |
| 3 PL LOG | banta | — | bántana | bántassi | bántappe |

^aLog, logophoric form.

(5a) ʔi ba šoor-úwa
 3.MASC.SING.SUBJ 3.LOG neighbor-MASC.ACC
 maadd-eési
 help.3.MASC.SING.IMPERF
 ‘he_x helps his_x neighbor’

(5b) ʔi ʔa
 3.MASC.SING.SUBJ 3.MASC.SING.POSS
 šoor-úwa maadd-eési
 neighbor-MASC.ACC help-3.MASC.SINGIMPERF
 ‘he_x helps his_y neighbor’

In the logophoric form (LOG), the gender distinction in the third-person singular form is neutralized.

Verbs

Subject Agreement, Aspect, Negation, and Modality

In affirmative declarative sentences, a three-way temporal distinction is made, for example, *beʔ-iísi* ‘he saw,’ *beʔ-eési* ‘he sees,’ *beʔ-aná* ‘he/she/I (etc.) will see.’ The verb shows subject agreement; object agreement is not marked on the verb (see **Tables 9** and **10**).

Future tense/aspect is formed by suffixing an invariable *-ána* to a verb root:

(6) kúnd-aná ‘I/you/he/she/we/you (PL)/
 they will see’

In negative declarative sentences, there is only two-way distinction, between perfective negative (**Table 11**) and imperfective negative (**Table 12**); the present and future forms are reduced to one paradigm: *beʔ-énna* ‘he does/will not see’ and *beʔ-ibénna* ‘he did not see.’

Interrogatives

There are the following content question words in Wolaitta: *ʔái* ‘what,’ *ʔái-gé* ‘which (MASC),’ *ʔái-nná* ‘which (FEM),’ *ʔái-ssi* ‘why,’ *ʔa-udé* ‘when,’ *ʔá-wan* ‘where,’ and *ʔoóni* ‘who.’

Table 9 Perfective paradigm

| Singular | Plural |
|-----------------------|----------------------------|
| kúnd-aási ‘I fell’ | kúnd-ida ‘we fell’ |
| kúnd-ádasa ‘you fell’ | kúnd-ideta ‘you (PL) fell’ |
| kúnd-iísi ‘he fell’ | kúnd-idosona ‘they fell’ |
| kúnd-aásu ‘she fell’ | |

Table 10 Present tense paradigm

| Singular | Plural |
|-----------------------|---------------------------|
| kúnd-aísi ‘I fall’ | kúnd-oósi ‘we fall’ |
| kúnd-aása ‘you fall’ | kúnd-eéta ‘you (PL) fall’ |
| kúnd-eési ‘he falls’ | kúnd-oósóna ‘they fall’ |
| kúnd-aúsu ‘she falls’ | |

Table 11 Perfective negative

| Singular | Plural |
|--------------------------------|---------------------------------------|
| beʔ-á-beíkke ‘I did not see’ | beʔ-i-boókkó ‘we did not see’ |
| beʔ-á-baákká ‘you did not see’ | beʔ-i-beékkétá ‘you (PL) did not see’ |
| beʔ-i-beénná ‘he did not see’ | beʔ-i-boókkóná ‘they did not see’ |
| beʔ-á-beíkkú ‘she did not see’ | |

Table 12 Imperfective negative

| Singular | Plural |
|----------------------------------|---------------------------------------|
| beʔ-íkke ‘I do/will not see’ | beʔ-ókkó ‘we do/will not see’ |
| beʔ-ákka ‘you do/will not see’ | beʔ-ékketa ‘you (PL) do/will not see’ |
| beʔ-énna ‘he does/will not see’ | beʔ-ókkóna ‘they do/will not see’ |
| beʔ-úkkú ‘she does/will not see’ | |

Table 13 Verbal inflection in interrogative sentences

| Interrogatives | | | |
|----------------|----------------------------------|---------------------------|----------------|
| Person | Perfective | Imperfective Pres/Hab. | Future |
| 1 SING | beʔ-ádina 'did I see?' | beʔ-aína | beʔ-ané |
| 2 SING | beʔ-ádi 'did you see?' | beʔ-áy | beʔ-uúte |
| 3 MASC SING | beʔ-ide 'did he see?' | beʔ-i | beʔ-ané |
| 3 FEM SING | beʔ-áde 'did she see?' | beʔ-áy | beʔ-ané |
| 1 PL | beʔ-ido 'did we see?' | beʔ-iyó | beʔ-ané |
| 2 PL | beʔ-ideti 'did you (PL) see?' | beʔ-eéti | beʔ- uútetí |
| 3 PL | beʔ-idona 'did they see?' | 'beʔ-iyona' | beʔ-ané |

Both in content-question-word and polar-interrogative clauses, the verb inflects for subject, tense/aspect, and modality (i.e., [+question]). The actual subject-agreement- and tense/aspect-marking morphemes are distinct from that observed for declarative sentences. Examples are shown in Table 13.

Imperative and Optative Moods

Second-person singular and plural imperatives are marked by *-á* and *-(i)ité*, respectively (Table 14). The optative/hortative involves only the third-person singular and plural forms. It is marked by *-ó* for third-person singular masculine and by *-ú* for feminine. For third-person plural, it is marked by *-óna*.

- (7a) demm-ó 'let him find'
 (7b) demm-ú 'let her find'
 (7c) demm-óna 'let them find'

The imperative and optative/hortative forms take the same negative marker, *-ópp-lúpp-*, which is formally distinct from the negation-marking morpheme in affirmative declarative sentences.

- (8a) demm-ópp-a 'don't find (2.SING)!'
 (8b) demm-ópp-ite 'don't find (2.PL)!'
 (9a) demm-ópp-ó 'let him not find!'
 (9b) demm-úpp-ú 'let her not find!'
 (9c) demm-ópp-óna 'let them not find'

Verb root extension in Wolaitta includes causative, passive, reciprocal, reflexive, and intensive verbs (Table 15).

Clauses

Simple Declarative Clauses

The most frequently used word order is SOV. However, S may occur immediately before V when it is in contrastive focus. Also, subject and object may be omitted.

Table 14 Imperative

| Singular | Plural | Gloss |
|----------------|---------------------|--------------------|
| demmm-á y-á | demmm-ité y-iité | 'find!' 'come!' |

Table 15 Verb root extensions

| Verb root | Causative stem | Passive/ reciprocal | Intensive/ repetitive | Gloss |
|-----------------|--------------------------------|-------------------------|-----------------------------|--------------------|
| k'ant'- bóg- | k'ant'- is(s)- bóg-is(s) | k'ant'-étt- bóg-étt- | k'ant'-erett- bóg-erett- | 'cut' 'plunder' |

- (10) ʔasa-t-í méhe-t-a
 person-PL-MASC.NOM cattle-PL-MASC.ABS
 baiz-idosona
 sell-3.PL.PERF
 'the people sold the cattle'

In phrases, modifiers precede the head. Demonstratives generally precede numerals and adjectives when both modify the same noun. Example (11a) is an NP with adjectives and a demonstrative; (11b) is a sentence containing an NP with a relative clause.

- (11a) ha heezú guútta naa-t-í
 this three small child-PL-MASC.NOM
 'these three small children'
- (11b) maay-úwa meec'c'-íya
 cloth-MASC.ACC wash-IMPERFEL
 naʔ-íya daapur-aásu
 child.FEM.NOM be_tired-3.FEM.SING.PERF
 'the girl who is washing clothes is tired'

Complex Clauses

In complex sentences, adverbial and complement clauses precede main clauses. Clausal linking is indicated by various verbal affixes attached to the dependent clause. Examples (12a) and (12b) are simultaneous, examples (12c) and (12d) are anterior and examples (12e) is conditional. Simultaneous and anterior morphemes further indicate whether the subject of the dependent clause is the same as that of the main clause.

- (12a) ʔattúma ʔasa-t-í keettaa
 male person-PL-NOM house.MASC.ACC
 keet't'-íšín mác'c'a ʔasa-t-í
 build-DS.SIMUL female person-PL-NOM
 puutt-úwa súk'k'-osona
 cotton-MASC.ACC spin-3.PL.IMPERF
 'when the men are building the house, the women
 spin cotton'

- (12b) ?attúma ?asa-t-í keettaa
male person-PL-NOM house.MASC.ACC
keet't'-íiddi /isso-y /iss-úwa
build-DS.SIMUL one-NOM one-ACC
k'ir-oósona
tease-3.PL.IMPERF
'the men tease each other while building the house'
- (12c) ?attúma ?asa-t-í keettaa
male person-PL-NOM house.MASC.ACC
keet't'-ín mac'c'a ?asa-t-í
build-DS.CNV female person-PL-NOM
gidd-úwa meeš-oósona
interior-MASC.ACC smear_dung-3.PL.IMPERF
'the men having built the house, the women smear
the interior with dung'
- (12d) ?asa-t-í keettaa keet't'-ídí
person-PL-NOM house.MASC.ACC build-SS.CNV
šemp-oósona
rest-3.PL.IMPERF
'the people rest having built the house'
- (12e) ?asa-t-í keettaa keet't'-íkko
male person-PL-NOM house.MASC.ACC
tá ?eta-w pars-úwa
rest-3.PL.IMPERF 3.PL.OBJ-DAT beer-MASC.ACC
?ag-ana
brew-FUT
'if the men build the house, I will brew them beer'

Bibliography

- Adams B A (1983). A tagmemic analysis of the Wolaitta language. Ph.D. diss., University of London.
- Azeb A (1996). 'Tone-accent and prosodic domains of Wolaitta.' *Studies in African Linguistics* 25(2), 111–138.
- Azeb A (2001). 'Ideophones and compound verbs in Wolaitta.' In Kilian-Hatz C & Voeltz F K E (eds.) *Ideophones*. Amsterdam, Philadelphia: John Benjamins. 49–62.
- Bekale S (1989). The case system in Wolayta. M.A. thesis, Addis Ababa University.
- Ethiopian Language Academy (1995). *Wolaitatto leemi-suwa. Wolaitta proverbs with Amharic translation*. Talachew G & Ammenu T (eds.) Addis Ababa, Ethiopia: Artistic Printers for Ethiopian Language Academy.
- Fleming H C (1976). 'Cushitic and Omotic.' In Bender M L *et al.* (eds.) *Language in Ethiopia*. London: Oxford University Press. 34–53.
- Lamberti M & Sottile R (1997). *The Wolayta language*. Köln: Köppe.
- Ohman W A, Fulass H, Keefer J, Keefer A, Taylor Ch V & Marcos H M (1976). 'Welamo.' In Bender M L *et al.* (eds.) *Language in Ethiopia*. London: Oxford University Press. 155–164.
- Yitbarek E (1983). The phonology of Wolaitta: generative approach. M.A. thesis, Addis Ababa University.

Wolof

F Mc Laughlin, University of Florida, Gainesville, FL, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Wolof is a member of the northern branch of the Atlantic family of Niger-Congo languages, formerly known as West Atlantic, and is spoken primarily in Senegal as well as in parts of Gambia and Mauritania on the West African coast. In Senegal, Wolof serves as a lingua franca, and is spoken by upwards of 80% of the population as either a first or second language, making for a total of no fewer than 6–7 million speakers and quite possibly more. Wolof society has traditionally been hierarchically stratified (Diop, 1981) and is composed of two main social groups, *ñeeño* and *géer*. The former group consists of endogamous artisans or castes, including griots (verbal artists), blacksmiths, leatherworkers, and musicians; the latter group is composed of noncasted people and

nobles. Today, a majority of Wolof speakers are Sufi Muslims, most having converted to Islam en masse in the late 19th and early 20th centuries.

Genetic Affiliation

Sapir (1971) hypothesized that Wolof, along with Serer-Sine and Pulaar or Fula, belongs to the Senegal subgroup of northern Atlantic languages. Although the three languages are clearly related, Serer-Sine and Pulaar resemble each other much more closely than either of them do Wolof. Until much more historical work is done on the northern Senegal languages, the exact relationship of Wolof to these languages, as well as to other Atlantic languages, and especially to the Cangin languages spoken around the Senegalese city of Thiès, will remain unresolved.

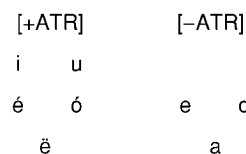
Phonetics and Phonology

Like most Niger-Congo languages (Clements, 2000), the consonantal inventory of Wolof, given in Table 1

Table 1 Wolof consonant articulation

| Consonant type | Labial | Alveolar | Palatal | Velar | Uvular |
|----------------|--------|----------|---------|-------|--------|
| Stops | p b | t d | c ɟ | k ɡ | q |
| Fricatives | f | s | x | | |
| Nasals | m | n | ɲ | ŋ | |
| Prenasalized | mb | nd | nɟ | ng | |
| Liquids | | l, r | | | |
| Glides | w | | y | | |

in standard Wolof orthography, distinguishes four main places of articulation: labial, alveolar, palatal, and velar. Voiced and voiceless stops, voiced prenasalized stops, voiceless fricatives, and simple nasal stops occur in the four places of articulation. Voiceless prenasalized stops no longer occur word-initially, but historical records and some place-names, such as Mpal, provide evidence that they once did. The have now been replaced by simple voiceless stops. There is also a voiceless uvular stop in the language, as in the words *sàq* ‘granary’ and *bëqët* ‘to be cowardly.’ There is a tap [ɾ] and a lateral [l] in addition to two glides, the labiovelar [w], and the palatal [j]. Consonant length is distinctive in Wolof: compare *dag* ‘valet’ to *dagg* ‘to cut,’ and *jaw* ‘to cook for a long time’ to *jaww* ‘sky’; however, not all consonants have a geminate counterpart, notably the prenasalized stops, the fricatives, and the alveolar tap. Geminate forms of the latter, however, occur in ideophones, as in *jérr* ‘of being hot’ and *curre* ‘of being red.’ Notable in the northern Atlantic context is the absence of implosive stops in Wolof. Wolof has an eight-vowel system in which vowels have either a plus or minus value for the advanced tongue root (ATR) feature. The [+ATR] vowels comprise the set i, u, é, ô, and ë; the [–ATR] vowels are e, o, and a (Figure 1). All vowels are written in standard Wolof orthography, and the character ë represents schwa. The [+ATR] and [–ATR] vowels are phonemically distinct in Wolof stems, as evidenced by the pairs *reer* ‘to dine’ and *réer* ‘to be lost,’ and *woor* ‘to fast’ and *wóor* ‘to be sure or trustworthy.’ Nominal and verbal stems and a substantial number of derivational suffixes harmonize for the ATR feature. Regressive height harmony also exists in the language. Vowel length is distinctive in Wolof, as in the pairs *bax* ‘to boil’ and *baax* ‘to be good’ and *fiit* ‘to tie on’ and *fiit* ‘soul,’ but the mid-central vowel ë does not have a long counterpart. Although most Niger-Congo languages are tonal, Wolof, like Serer-Sine and Pulaar, is not a tonal language. Intonational patterns are fairly flat according to Rialland and Robert (2001), and stress falls on the initial syllable of a word in Wolof.

**Figure 1** The Wolof eight-vowel system; vowels have either a plus or minus value for the advanced tongue root (ATR).

Morphology and Syntax

Wolof has a noun class system comprising 10 classes, of which 8 are singular and 2 are plural, marked by a single consonant. Unusually, there is no morphological marking for class on the noun, but the classifier consonant appears on nominal determiners as in the following examples, in which the determiner follows the noun:

1. m-class: *picc mi* ‘the bird,’ *picc male* ‘that bird.’
2. y-class: *picc yi* ‘the birds,’ *picc yale* ‘those birds.’
3. k-class: *nit ki* ‘the person,’ *nit kale* ‘that person.’
4. ñ-class: *nit ñi* ‘the people,’ *nit ñale* ‘those people.’

Wolof has approximately 30 verbal extensions, inflectional and derivational affixes that encode a variety of concepts such as reciprocal, applicative, causative, locative, etc. The verb *gis* ‘to see’ has, among others, the following derivatives: *gis* ‘to see,’ *gisaa* ‘to see again,’ *gise*, *gisante* ‘to see each other,’ *gisandoo* ‘to see together,’ and *gisaa* ‘to see (‘while you’re at it).’ Verb-to-noun derivation may exhibit reduplication (*gis* ‘to see,’ *gis-gis* ‘opinion’; *xam* ‘to know,’ *xam-xam* ‘knowledge’), suffixation (*gudd* ‘to be long,’ *guddaay* ‘length’), and consonant mutation (*baax* ‘to be good,’ *mbaax* ‘goodness’; *sonn* ‘to be tired,’ *cono* ‘fatigue’). It is arguable as to whether a distinct category of adjectives can be said to exist in Wolof, since adjectival forms can be subsumed under the category of verb (Creissels, 2000; Mc Laughlin, 2004). Although basic word order in Wolof is subject-verb-object, the information structure of Wolof is encoded in an elaborate focus system (Creissels and Robert, 1998). The minimal verb phrase consists of a bare verb plus an auxiliary that encodes person, number, and focus. Examples (1)–(4) show four different ways to say ‘Ami saw the thief,’ using neutral, subject, object, and verbal focus, respectively:

- (1) Ami gis na sàcc ba.
 Ami see 3S:PERF thief DET
- (2) Ami moo gis sàcc ba.
 Ami 3S:SFOC see thief DET
- (3) Sàcc ba la Ami gis.
 Thief DET 3S:OFOC Ami see
- (4) Ami dafa gis sàcc ba.
 Ami 3S:VFOC see thief DET

Urban Wolof

Urban Wolof, and especially that of the capital, Dakar, exhibits heavy lexical borrowing from French, as in Examples (5) and (6) (Mc Laughlin, 2001) (French loans are in boldface):

(5) Feu bi rouge na.
light DET be red 3S:PERF
'The traffic light turned red.'

(6) Dafa d-oon **errer** ci
3S:VFOC IMPERF-PAST wander PREP
monde bi rekk.
world DET just
'He was just wandering around the world.'

Bibliography

- Clements G N (2000). 'Phonology.' In Heine B & Nurse D (eds.) *African languages: an introduction*. Cambridge: Cambridge University Press. 123–160.
- Creissels D (2003). 'Adjectifs et adverbes dans les langues subsahariennes.' In Sauzet P & Zribi-Hertz A (eds.) *Typologie des langues d'Afrique et universaux de la grammaire*, vol. 1. Paris: L'Harmattan. 17–38.
- Creissels D & Robert S (1998). 'Morphologie verbale et organization discursive de l'énoncé: l'exemple du tswana et du wolof.' *Faits de Langues* 11–12, 161–178.
- Diop A-B (1981). *La société wolof: les systèmes d'inégalité et de domination*. Paris: Karthala.
- Diouf J-L (2001). *Grammaire du wolof contemporain*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Diouf J-L (2003). *Dictionnaire wolof-français et français-wolof*. Paris: Karthala.
- Ka O (1994). *Wolof phonology and morphology*. Lanham, MD: University Press of America.
- Kihm A (1999). 'Focus in Wolof: a study of what morphology may do to syntax.' In Rebuschi G & Tuller L (eds.) *The grammar of focus*. Amsterdam: John Benjamins. 245–273.
- Mc Laughlin F (1997). 'Noun classification in Wolof: when affixes are not renewed.' *Studies in African Linguistics* 26, 1–28.
- Mc Laughlin F (2001). 'Dakar Wolof and the configuration of an urban identity.' *Journal of African Cultural Studies* 14, 153–172.
- Mc Laughlin F (2004). 'Is there an adjective class in Wolof?' In Dixon R M W & Aikhenvald A Y (eds.) *Adjective classes: a cross-linguistic typology*. Oxford: Oxford University Press. 242–262.
- Njie C M (1982). *Description syntaxique du wolof de Gambie*. Dakar, Abidjan & Lomé: Les Nouvelles Éditions Africaines.
- Pulleyblank D (1996). 'Neutral vowels in optimality theory: a comparison of Yoruba and Wolof.' *Canadian Journal of Linguistics* 24, 295–347.
- Rialland A & Robert S (2001). 'The intonational system of Wolof.' *Linguistics* 39, 893–939.
- Robert S (1991). *Approche énonciative du système verbal: le cas du wolof*. Paris: Éditions du CNRS.
- Sapir J D (1971). 'West Atlantic: an inventory of the languages, their noun class systems and consonant alternation.' In Sebeok T (ed.) *Current trends in linguistics 7: linguistics in sub-Saharan Africa*. The Hague: Mouton. 45–112.



Xhosa

M W Visser, University of Stellenbosch, Stellenbosch, South Africa

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Xhosa (or isiXhosa, with the noun class prefix) belongs, with isiZulu (Zulu) and isiNdebele (Ndebele), to the Zunda subgroup of the Nguni group of the Southeastern Zone of Bantu languages. This zone also includes the Sotho, Venda, and Tsonga language groups. In terms of Guthrie's (1967–1971) classification, isiXhosa is identified as S41 (Doke, 1943, 1954; Piron, 1998; Gowlett, 2003; Nurse and Philippson, 2003). The Bantu family forms part of the larger Niger-Congo family of African languages of which the three other major families are Afroasiatic, Nilo-Saharan, and Khoisan (Greenberg, 1963; Heine and Nurse, 2000b; Williamson and Blench, 2000). Specific areas in the Eastern Cape province of South African have historically been associated with the various dialects or local forms of isiXhosa, namely isiGcaleka, isiNdlambe, isiGaika, isiThembu, isiBomvana, isiMpondomise, isiMpondo, and isiXesibe.

With the establishment of a democratic South Africa in 1994, isiXhosa has obtained the status of an official language, together with eight other Bantu languages spoken in South Africa, namely isiZulu, isiNdebele, Siswati (Swati), Sesotho (Southern Sotho), Sepedi/Sesotho sa Leboa (Northern Sotho), Setswana (Tswana), Tshivenda (Venda), and Xitsonga (Tonga). The government has introduced significant legislation through the Department of Arts and Culture for promoting the status and use of these official languages in government, education, and business, in addition to the predominant use of English. Huge challenges exist for accomplishing this goal, which includes urgent work in the fields of terminology development, language in education policy, and the teaching and learning of the indigenous African languages (Webb, 2001; Visser, 2004, 2005).

Nouns and Noun Phrases

Noun Classes

The morphology and semantics of the noun class system of isiXhosa is typical of the Bantu languages (Greenberg, 1963; Welmers, 1973; Du Plessis, 1978; Poulos and Msimang 1998; Piron 1998; Williamson and Blench, 2000; Gowlett, 2003). IsiXhosa has nouns in all noun classes from class 1 to 15, excluding class 12 and 13. The locative classes 16, 17, and 18 are morphologically fossilized; thus, they all exhibit the associated locative agreement subject- and object-verb agreement morpheme *ku-* rather than the distinct agreement morphemes of class 16, 17, and 18. As is general to the Bantu languages for the first 10 classes, the consecutive odd and even class numbers are regular singular-plural pairs. The noun class prefixes have a VCV syllable structure, except for classes 1, 3, and 9; the postnasal vowel in classes 1 and 3 has been deleted, and class 9 has the prefix *in-*. Classes 1a and 2a, subclasses of classes 1 and 2, respectively, have only vowel prefixes. **Table 1** shows the noun class prefixes of isiXhosa.

Table 1 Noun class prefixes of IsiXhosa

| Noun class | Prefix | Example noun |
|------------|-----------|-------------------------------------|
| 1 | um- | umfazi 'woman' |
| 2 | aba- | abafazi 'women' |
| 1a | u- | utata 'father' |
| 2a | oo- | ootala 'fathers/father and company' |
| 3 | um- | umlilo 'fire' |
| 4 | imi- | imililo 'fires' |
| 5 | i(li)- | ilitye 'stone' |
| 6 | ama- | amatye 'stones' |
| 7 | isi- | isiqhamo 'fruit' |
| 8 | izi- | iziqhamo 'fruits' |
| 9 | i(n)- | indlu 'house' |
| 10 | i(z)i(n)- | izindlu 'houses' |
| 11 | ulu- | uluthi 'stick' |
| 14 | ubu- | ubusika 'winter' |
| 15 | uku- | ukutya 'food' |

Table 2 Nominal suffixes: feminine, augmentative, and diminutive

| | | |
|-----------|--|--|
| -kazi FEM | inkosi 'chief'; ixhego 'old man'; utitshala 'teacher'; | inkosikazi 'chieftainness' ixhegokazi 'old woman' utitshalakazi 'female teacher' |
| -kazi AUG | umthi 'tree'; intaba 'mountain'; indlu 'house' | umthikazi 'big tree' intabakazi 'big mountain' indlukazi 'big house' |
| -ana DIM | indoda 'man'; incwadi 'book'; ilitye 'stone'; | indodana 'small man' incwadana 'small book' ilityana 'small stone' |

Source: Du Plessis (1978, 1997); Louw (1963).

Nominal Suffixes

Nouns in isiXhosa can regularly take suffixes that denote the property of feminine *-azi-*, augmentative *-azi*, and reciprocal *-ana*, as shown in Table 2.

Agreement Morphology with Nominal Modifiers

As is characteristic of the Bantu languages, isiXhosa exhibits agreement morphology of the nominal modifiers with the head noun, where the latter may be a lexical noun or a phonetically empty pronominal (Doke, 1954; Greenberg, 1963; Guthrie, 1967–1971; Welmers, 1973; Du Plessis and Visser, 1992; Gowlet 2003; Nurse and Philippson, 2003b). The nominal modifiers identified for isiXhosa include demonstratives, adjectives, nominal relatives, clausal relatives, numerals, quantifiers, possessives, and enumeratives (Louw, 1963; Du Plessis, 1978, 1983; Visser, 1984, 2002; Du Plessis and Visser, 1992). The examples that follow illustrate the agreement morphology of the adjective and possessive with pairs of lexical head nouns in classes 1, 2, 5, 6, 7, and 8.

- The head noun is in class 1.

| | | | |
|------|----------------------|--------------|---------------|
| (1a) | umntwana | wam | omhle |
| | umntwana | u-a-m | om-hle |
| | child | AGR-GEN-mine | AGR-beautiful |
| | 'my beautiful child' | | |

- The head noun is in class 2.

| | | | |
|------|-------------------------|--------------|---------|
| (1b) | abantwana | bam | abahle |
| | abantwana | ba-a-m | aba-hle |
| | children | AGR-GEN-mine | AGR-hle |
| | 'my beautiful children' | | |

- The head noun is in class 5.

| | | | |
|------|----------------------|--------------|---------------|
| (1c) | ihashe | lam | elihle |
| | ihashe | li-a-m | eli-hle |
| | horse | AGR-GEN-mine | AGR-beautiful |
| | 'my beautiful horse' | | |

Table 3 Deverbal nouns

| Verb | Deverbal noun | |
|------------------------------------|--|--|
| | Human | Nonhuman |
| -thenga 'buy' | umthengi 'buyer' (class 1) | intengo 'buy' (class 9) |
| -funda 'read, learn' | umfundi 'learner, student' (class 1) | imfundo 'education' (class 9) |
| -dlala 'play' | umdlali 'player' (class 1) | umdlalo 'game' (class 3) |
| -hamba 'travel', 'gula 'be ill' | umhambi 'traveller' (class 1) isigulana 'patient' (class 7) | uhambo 'travel' (class 11) ingulo 'illness' (class 9) |
| -thanda 'like, love' | isithandwa 'beloved' (class 7) | uthando 'love' (class 11) |

- The head noun is in class 6.

| | | | |
|------|-----------------------|--------------|---------------|
| (1d) | amahashe | am | amahle |
| | amahashe | a-a-m | ama-hle |
| | horses | AGR-GEN-mine | AGR-beautiful |
| | 'my beautiful horses' | | |

- The head noun is in class 7.

| | | | |
|------|---------------------|-----------|---------|
| (1e) | isitya | sam | esihle |
| | isitya | si-a-m | esi-hle |
| | dish | AGR-GEN-m | AGR-hle |
| | 'my beautiful dish' | | |

- The head noun is in class 8.

| | | | |
|------|-----------------------|--------------|---------------|
| (1f) | izitya | zam | ezihle |
| | izitya | zi-a-m | ezi-hle |
| | dishes | AGR-GEN-mine | AGR-beautiful |
| | 'my beautiful dishes' | | |

Derived Nouns IsiXhosa exhibits regular nominal derivation from verbs and, to a less regular degree, from other word categories such as adjectives and nominal relatives (Louw, 1963; Du Plessis, 1978). The examples in Table 3 illustrate deverbal nouns in a range of noun classes.

Compound Nouns

Compound nouns are common in isiXhosa, and this is especially salient in proper nouns.

| | | | |
|------|-----------------------|---|-------------------|
| (2a) | umninikhaya | < | umnini-ikhaya |
| | 'home owner' | | owner-home |
| (2b) | impilontle | < | impilo-entle |
| | 'good health' | | life-AGR-good |
| (2c) | indlalifa | < | indla-ilifa |
| | 'person who inherits' | | eater-inheritance |
| (2d) | imalimboleko | < | imali-imboleko |
| | 'loan' | | money-loan |

- (2e) uNtombizobawo < u-ntombi-za-ubawo
'girls of father' AGR(cl.1)-girls-of-father'
- (2f) uNoxolo < u-No-uxolo
'the one with peace' AGR(cl.1)-Fem-peace
- (2g) uMzimkhulu < u-Mzi-M-khulu
'big house' AGR(cl.1)-house-AGR-big
- (2h) uNtombi zandile < u-Ntombi-zi-and-ile
'the girls have increased' AGR(cl.1)-girls-AGR(cl.10)-increase-Perf.

Verbs, Verb Phrases, and Clauses

Transitivity and Verbal Derivation

IsiXhosa has a wide range of nonderived verbs, which are intransitive and monotransitive. A smaller number of nonderived verbs are ditransitive, as shown in the following examples. Intransitive verbs (3) include experienter verbs, motion verbs, and weather verbs.

- (3a) -gula 'be ill'
(3b) -vuya 'be happy'
(3c) -sebenza 'work'
(3d) -hamba 'travel'
(3e) -phuma 'go out, exit'
(3f) -tshona 'sink'
(3g) -jika 'turn'
(3h) -buya 'return'

Verbs from a wide range of semantic classes appear as nonderived monotransitive verbs, as illustrated by the following examples.

- Verbs of change.

- (4a) -aphula 'break'
(4b) -goba 'bend'
(4c) -pheka 'cook'
(4d) -oja 'roast'
(4e) -vala 'close'

- Verbs of change of possession.

- (5a) -qokelela 'collect'
(5a) -fumana 'get, obtain'
(5a) -(i)ba 'steal'
(5a) -kha 'pick (fruit)'

- Verbs of communication.

- (6a) -bika 'report'
(6b) -thetha 'speak'
(6c) -ncokola 'converse'
(6d) -hleba 'gossip'
(6e) -geza 'joke'

- Verbs of contact.

- (7a) -beka 'put'
(7b) -tyala 'plant'
(7c) -xhoma 'hang'

- (7d) -sula 'wipe'
(7e) -khupha 'take out'
(7f) -galela 'pour'

- Verbs of creation.

- (8a) -qingqa 'carve'
(8b) -xovula 'knead'
(8c) -zoba 'draw'
(8d) -akha 'build'
(8e) -bhaka 'bake'

- Verbs of perception.

- (9a) -bona 'see'
(9b) -(i)va 'hear/feel/taste'
(9c) -ngcamla 'taste'
(9d) -ngqina 'witness'
(9e) -jonga 'look at'

- Verbs of social interaction.

- (10a) -vuma 'agree'
(10a) -qhula 'joke'
(10a) -tyelela 'visit'

Examples of nonderived ditransitive verbs appear in (11).

- (11a) -nika 'give'
(11b) -pha 'give (as gift)'
(11c) -boleka 'lend'
(11d) -vimba 'refuse'
(11e) -buza 'ask'
(11f) -cela 'request'

As is typical of the Bantu languages, the transitivity properties of verbs in isiXhosa can be altered by suffixation of various verbal derivational suffixes, which can appear in combination with one another (Satyo, 1986) and which can be reduplicated to achieve various semantic effects. The applicative (APPLIC) and causative suffixes are transitivizing suffixes in that they introduce a new NP argument to the verb (Du Plessis, 1978, 1980b, 1997; Du Plessis and Visser 1992, 1998). When these suffixes appear, intransitive verbs become monotransitive and monotransitive verb become ditransitive. The applicative suffix can introduce an NP argument bearing the semantic role of benefactive, malefactive, recipient, purpose, and cause/reason, as shown in the examples in (12) (AGRS stands for subject agreement).

- (12a) umfazi ufundela
 umfazi u-fund-el-a
 woman AGRS-read-APPLIC-PRES
 abantwana amabali
 abantwana amabali
 children stories
 'the woman reads stories for the children'

Table 4 Reduplicated applicative forms

| <i>Verb</i> | <i>Reduplicated applicative</i> |
|----------------|---------------------------------|
| -bopha 'tie' | -bophelela 'tie thoroughly' |
| -funa 'search' | -funelela 'search thoroughly' |
| -sula 'wipe' | -sulelela 'wipe thoroughly' |

- (12b) inkwenkwe ibalekela indebe
 inkwenkwe i-balek-el-a indebe
 boy AGRS-run-APPLIC-PRES cup
 'the boy is running for the cup'

- (12c) umfazi ulilela ilahleko
 umfazi u-lil-el-a ilahleko
 woman AGRS-cry-APPLIC-PRES loss
 'the woman cries for (her) loss'

- (12d) abafazi baphekela
 abafazi ba-pheke-el-a
 women AGRS-cook-APPLIC-PRES
 umtshato inyama
 umtshato inyama
 wedding meat
 'The women cooks meat for the wedding'

The applicative can appear in a reduplicated form to denote an intensified action, as shown in Table 4.

Causative Suffix

The causative (CAUS) suffix *-is-* regularly denotes three kinds of meanings, depending on the verbal semantics and the pragmatic context: coercive ('make/force to do something'), assistive ('help do something'), and permissive ('let do something').

- (13a) umfana ulimisa utata intsimi
 umfana u-lim-is-a utata intsimi
 young.man AGRS-plough- father field
 CAUS-PRES

'the young man helps his father plough the field'

- (13b) utitshala ubhalisa abantwana ileta
 utitshala u-bhal-is-a abantwana ileta
 teacher AGRS-write- children letter
 CAUS-PRES

'the teacher makes/helps/lets write the children a letter'

Detransitivising Verbal Affixes

The reciprocal (RECIP) suffix *-ana* (14) and the reflexive (REFL) verbal prefix *-zi-* (15) in isiXhosa are detransitivizing morphemes, as is typical of the Bantu languages.

- (14a) abantu bayathandana
 abantu ba-ya-thand-an-a
 people AGRS-PRES-like-RECIP-PRES
 'the people like each other'

- (14b) uZola noNomsa bayathandana
 uZola na-uNomsa AGRS-PRES-thand-an-a
 Zola and-Nomsa like-RECIP-PRES
 'Zola and Nomsa like/love each other'

- (14c) uZola uthandana noNomsa
 uZola u-thand-an-a na-uNomsa
 Zola AGRS-like-RECIP-PRES with-Nomsa
 'Zola and Nomsa like/love each other'

- (15a) umntwana uyazibona
 umntwana u-ya-zi-bon-a
 child AGRS-PRES-REFL-see-PRES
 'the child sees himself/herself'

- (15b) abantwana bayazihlamba
 abantwana ba-ya-zi-hlamb-a
 children AGRS-PRES-REFL-wash-PRES
 'the children were themselves'

Unaccusative Verbal Suffixes: Passive and Neuter-Passive

Passive (PASS) *-w-* (16) and neuter-passive (NEUT. PASS) stative *-ele-/akal-* (17) verbal suffixes are unaccusative, in that the object of a transitive verb must either raise to become the subject of the verb or remain in the object position and receive nominative case from a phonetically empty existential subject pronominal associated with the subject agreement prefix *ku-* on the verb (Visser, 1986; Du Plessis and Visser, 1992b, 1998).

- (16a) incwadi iyafunwa ngumfundi
 incwadi i-ya-fun-w-a ng-umfundi
 book AGRS-PRES- want-PASS-pres
 'the book is wanted/searched for by the student'

- (16b) kufunwa incwadi ngumfundi
 ku-fun-wa incwadi ng-umfundi
 EXIST.AGRS-want/ book cop-student
 search-PASS.PRES
 'there is being wanted a book by the student'

- (17a) incwadi iyafuneka kumfundi
 incwadi i-ya-fun-ek-a ku-umfundi
 book AGRS-PRES-want to-student
 -NEUT.PASS-PRES
 'a book is needed to (for) the student'

- (17b) kufuneka incwadi kumfundi
 ku-fun-ek-a incwadi ku-umfundi
 EXIST.AGRS-want book to-student
 -NEUT.PASS-PRES
 'there is a book is needed to (for) the student'

- (17c) intaba iyabonakala
 intaba i-ya-bon-akal-a
 mountain AGRS-PRES-see-NEUT.PASS-PRES
 'the mountain is visible'

- (17d) kubonakala intaba
 ku-bon-akal-a intaba
 EXIST.AGRS-see-NEUT.PASS-PRES mountain
 'the mountain is visible'

Table 5 Intransitive-transitive verbal pairs with the consonant *-k-* and *-l-*

| <i>Intransitive stative</i> | <i>Transitive</i> |
|-------------------------------|------------------------|
| -guquka 'be turned' | -guqula 'turn' |
| -aphuka 'be broken' | -aphula 'break' |
| -ahluka 'be separated/parted' | -ahlula 'separate' |
| -phekula 'be turned upside' | -phekula 'turn upside' |
| -khawuka 'be broken off' | -khawula 'break off' |
| -sombuluka 'be unfolded' | -sombulula 'unfold' |

Source: Du Plessis and Visser (1998).

The neuter-passive suffix *-ek-/akal-* changes the verb into a stative verb, as shown in Table 5.

Verbal Inflection

IsiXhosa exhibits the inflectional morphemes typical of the Bantu languages, namely agreement, tense, aspect, mood, and negative (Du Plessis, 1986, 1997; Du Plessis and Visser, 1992b, 1998; Gowlett, 2003).

Subject and Object Agreement Prefixes The isiXhosa verb, like verbs in the Bantu languages in general, exhibits a subject agreement prefix (AGRS), which appears obligatorily, except with imperative mood verbs and certain instances of deficient verbs. The isiXhosa verb also contains an object agreement prefix (AGRO), which in general appears optionally (Du Plessis, 1978, 1997) and is often used to emphasize the verb phrase or to establish the object feature when the object argument is separated from the verb by intervening lexical or phrasal categories. In the examples in (18), the noun classes of the NP subject and object appear in brackets.

- (18a) umntwana uyayifunda incwadi
 umntwana u-ya-yi-fund-a incwadi
 [class 1] [class 9]
 child AGRS-PRES- AGRO-read-PRES
 'the child reads a book'
- (18b) abantwana bayazifunda iincwadi
 abantwana ba-ya-zi-funda iincwadi
 [class 2] [class 10]
 children AGRS-PRES-AGRO- read.PRES
 'the children read the books'
- (18c) amadoda ayabusela utywala
 amadoda a-ya-bu-sela utywala
 [class 6] [class 14]
 men AGRS-PRES- AGRO-drink
 'the men drink beer'

Sentences like these, in which the object co-occurs with an object agreement prefix, denote additional emphasis on the verb phrase.

Aspect Morphemes The verbal inflectional morphology in isiXhosa contain a number of prefixes that denote aspectual features. These prefixes include *-sa-* 'still' (the progressive, PROG), *-ka-* 'not, get', and *-yawa-* 'as usual'. The potential morpheme *-nga-* denotes 'ability', 'possibility', or 'permission' (Louw, 1963; Du Plessis, 1978, 1997).

- (19a) abafundi basafunda
 abafundi ba-sa-fund-a
 students AGRS-PROG-learn/read-PRES
 'the students are still reading/learning'
- (19b) abafundi abakafundi
 abafundi a-ba-ka-fund-i
 students NEG-AGRS-not-read-NEG
 'the students have not read yet'
- (19c) abafundi bayawafunda
 abafundi ba-yawa-fund-a
 students AGRS-as.usual-read-PRES
 'the students are studying as usual'
- (19d) abafundi bangaphumelela
 abafundi ba-nga-phumelela
 student AGRS-can/may-succeed
 'the students can/may succeed'

Tense Inflection IsiXhosa has the typical tense distinctions found in the Bantu languages: present tense, perfect past tense, remote (A-) past tense, future tense, and compound (recent and remote) past tenses. The compound tenses appear as complex sentence constructions with a deficient verb taking a participial clause complement. The various past tenses are associated with specific features of (im)perfectivity (Louw, 1963; Du Plessis, 1978, 1986, 1997; Poulos and Msimang, 1998).

Present Tense The present tense verb form in isiXhosa can exhibit features of habituality and emphasis, in addition to denoting a literal present tense activity (Du Plessis, 1986, 1997).

- (20a) iintombi ziphendula imibuzo
 iintombi zi-phendul-a imibuzo
 girls AGRS-answer-PRES questions
 'the girls answer the questions'
- (20b) iintombi aziphenduli mibuzo
 iintombi a-zi-phendul-i mibuzo
 girls NEG-AGRS-answer- NEG questions
 'the girls do not answer (any) questions'
- (20c) iintombi aziyiphenduli imibuzo
 iintombi a-zi-yi-phendul-i imibuzo
 girls NEG-AGRS-AGRO -answer-NEG questions
 'the girls do not answer the (specific) questions'

The negative sentences in (20b) and (20c) illustrate the indefinite and definite negative, respectively. The

former is characterized by the absence of the initial vowel (the preprefix) of the object NP and the related absence of the object agreement prefix in the verb morphology, whereas the latter is characterized by the presence of the preprefix of the NP object argument and the associated presence of the object agreement prefix in the verb morphology (Visser, 2002). Similar definite and indefinite negatives may appear in all the other tenses.

Future Tense The future tense is characterized by the verb *-za* 'come' or *-ya* 'go' followed by the infinitive prefix on the main verb.

- (21a) iintombi ziza/ziya
 iintombi zi-za/zi-ya
 girls AGRS-come/AGRS-go
 kuphendula imibuzo
 ku-phendula imibuzo
 INF-answer questions
 'the girls will answer the questions'

The use of the deficient verb *-za* in the future tense denotes an immediate future, whereas the use of the deficient verb *-ya* denotes either a remote future or an immediate future action with a high degree of certainty.

- (21b) iintombi azizi/aziyi
 iintombi a-zi-z-i/a-zi-y-i
 girls NEG-AGRS-come-NEG/NEG-AGRS-go-NEG
 kuphendula mibuzo
 ku-phendula mibuzo
 INF-answer questions
 'the girls will answer the questions'

Perfect Past Tense The perfect past tense denotes an action in the recent past that has been completed (Louw, 1963; Du Plessis, 1978, 1997).

- (22a) abafana basebenzile
 abafana ba-sebenz-ile
 young.men AGRS-work-PERF
 'the young men worked'
 (22b) abafana abasebenzanga
 abafana a-ba-sebenz-anga
 young.men NEG-AGRS-work-PERF.NEG
 'the young men did not work'

Remote Past Tense The remote past tense verb takes a subject agreement prefix with a long rising-falling vowel *-a-*.

- (23a) iintombi zacula iingoma
 iintombi zi-a-cula iingoma
 girls AGRS-PAST-sing songs
 'the girls sang songs'

- (23b) umfundi wabhala ileta
 umfundi u-a-bhala ileta
 student AGRS-PAST-write letter
 'the student wrote a letter'

Compound Past Tenses The compound past tenses denote an activity or state that took place in the past but that has not been completed; hence, they exhibit the imperfective aspect. The lexical verb in these tenses are in the participial mood.

Recent Compound Past Tense The recent compound past tense is characterized by a perfect tense deficient verb *-be* taking a participial complement clause, as shown in the following examples.

- (24a) iintombi zibe zacula iingoma
 iintombi zi-b-e zi-cula iingoma
 girls AGRS-be-PERF AGRS-sing songs
 'the girls were singing songs'

- (24b) iintombi zibe zingaculi ngoma
 iintombi zi-b-e zi-nga-cul-i ngoma
 girls AGRS-be AGRS-NEG songs
 -PERF -sing-NEG
 'the girls were not singing (any) songs'

Remote Compound Past Tense The deficient verb *-ba/-ye* appears in the remote compound past tense taking the morpheme *-a-* in its subject agreement affix and subcategorizing for a participial complement clause, as shown in the following examples.

- (25a) iintombi zaba zacula iingoma
 iintombi zi-a-ba zi-cula iingoma
 girls AGRS-PAST-be AGRS-sing songs
 'the girls were singing songs'
 (25b) iintombi zaba zingaculi ngoma
 iintombi zi-a-ba zi-nga-cul-i ngoma
 girls AGRS- AGRS-NEG- songs
 PAST-be sing-NEG
 'the girls did not sing (any) songs'

Negative Inflection The negative inflection of isiXhosa is realized through verbal prefixation, infixation, and suffixation, depending on the mood properties of the verb (Du Plessis, 1978, 1997). The examples of negative sentences given in the previous section demonstrate that negation in indicative mood verbs is realized by a verbal prefix that occurs before the subject agreement prefix and by a verbal suffix, whereas in participial clauses (in the compound tenses) the negative morpheme (*-nga-*) appears after the subject agreement prefix together with a negative verbal suffix. Further examples of negative sentences in isiXhosa appear in the subsection on mood inflection next.

Mood Inflection Linguists differ on the number of moods that can be distinguished for isiXhosa and closely related languages such as isiZulu (Louw, 1963; Du Plessis, 1978, 1997; Poulos and Msimang, 1998). The following nine moods have been distinguished for isiXhosa.

Indicative Mood The indicative mood is used in main clauses for statements and questions. Indicative mood clauses may also appear as a complement clauses of factive verbs.

- (26a) abafundi bahala uviwo
 abafundi ba-bhal-a uviwo
 student AGRS-write-PRES examination
 'the student is writing examination'
- (26b) abafundi ababhali viwo
 abafundi a-ba-bhal-i viwo
 students NEG-AGRS examination
 -write-NEG
 'the students are not writing (any)
 examination'
- (26c) abafundi abalubhali uviwo
 abafundi a-ba-lu-bhal-i uviwo
 students NEG-AGRS-AGRO- examination
 write-NEG
 'the students are not writing the (specific)
 examination'

The sentences in (26) can be changed into questions by using rising intonation toward the sentence-final position.

The sentences in (26b) and (26c) illustrate the indefinite and definite negatives, respectively. The indefinite negative is characterized by the loss of the initial vowel of the noun class prefix of the object noun and the absence of the object agreement affix in the verbal morphology. The definite negative is characterized by the presence of the initial vowel of the object noun and the associated object agreement prefix in the verbal morphology. The indefinite-definite negative distinction also has an influence on the morphological form of several categories that function as nominal modifiers.

The indicative mood exhibits the tense distinctions discussed in the previous subsection on tense inflection.

Participial (Situative) Mood The participial mood is used in subordinate clauses that denote an activity or state that takes place simultaneously with the activity or state expressed by the main clause. It is clearly identifiable by its subject agreement morphology for noun classes 1, 2, and 6 and by morphemes that occur with monosyllabic and vowel verb stems in positive

sentences. In addition, the participial mood regularly occurs after certain temporal conjunctives (as in 27c) and deficient verbs (as in 27d) (Louw, 1963; Du Plessis, 1978, 1997; Du Plessis and Visser, 1992b).

- (27a) abafundi bathula
 abafundi ba-thula
 students AGRS-be.quiet
 bebhala uviwo
 be-bhala uviwo
 AGR(PART)-write examination
 'the students are quiet while they are writing
 examination'
- (27b) abafundi babhala uviwo
 abafundi ba-bhala uviwo
 students AGRS-write examination
 besiva imiyalelo
 be-si-v-a imiyalelo
 AGRS(PART)-AFF(PART)-hear-PRES instructions
 'the students write examinations while hearing
 the instructions'
- (27c) abafundi bathula xa
 abafundi ba-thula xa
 students AGRS-be.quiet when
 bebhala uviwo
 be-bhal-a uviwo
 AGRS(PART)-write-PRES examination
 'the students are quiet when they write
 examinations'
- (27d) abafundi basoloko
 abafundi ba-soloko
 students AGRS-always.do
 bebhala uviwo
 be-bhal-a uviwo
 AGRS(PART)-write-PRES examination
 'the students always write examination'
- (27e) abafundi babhala uviwo
 abafundi ba-bhal-a uviwo
 students AGRS-write-PRES examination
 bengafundanga kakuhle
 be-nga-fund-anga kakuhle
 AGRS(PART)-NEG-learn-NEG well
 'the students write examinations (while) they
 have not studied well'

The participial mood in isiXhosa can exhibit all the various tense forms discussed in the subsection on tense inflection.

Relative Mood The relative mood clause occurs widely as a nominal modifier in isiXhosa. It is characterized by the coalescence of a definitizing morpheme *-a-*, which also occurs with various other nominal modifiers, with the subject agreement prefix of the relative clause verb. This definitizing morpheme is, however, omitted when the relative clause head is the object argument of an indefinite negative verb or when the relative clause head occurs with

a demonstrative pronoun. The relative clause in isiXhosa typically contains a resumptive pronoun, coreferential with the relative clause head, which can be realized as an object agreement prefix in the verbal morphology, a prepositional complement, or complement of a copulative, as illustrated by the following examples:

- (28a) abafundi ababhala uviwo
 abafundi a-ba-bhal-a uviwo
 students AFF(DEF)-AGRS- examination
 write-PRES
 bafunde kakuhle
 ba-fund-e kakuhle
 AGRS-learn-PERF well
 'the students who write examinations have studied well'
- (28b) aba bafundi babhala
 aba bafundi ba-bhal-a
 DEM students AGRS-write-PRES
 uviwo bafunde kakuhle
 uviwo ba-fund-e kakuhle
 examinations AGRS-learn-PERF well
 'these students who write examinations have studied well'
- (28c) asibizi bafundi
 a-si-biz-i bafundi
 NEG-AGRS-call-NEG students
 babhala uviwo
 ba-bhal-a uviwo
 AGRS-write-PRES examinations
 'we do not call (any) students who write examinations'
- (28d) abafundi enibafunayo
 abafundi a-ni-ba-fun-a-yo
 students AFF(DEF)-AGRS-AGRO-want-PRES-AFF(RC)
 babhala uviwo
 ba-bhal-a uviwo
 AGRS-write-PRES examination
 'the students who you want, are writing examination'
- (28e) abafundi endiya kubo
 abafundi a-ndi-y-a ku-bo
 students AFF(RC)-I-go-PRES to-them
 babhala uviwo
 ba-bhal-a uviwo
 AGRS-write-PRES examination
 'the students to whom I am going write examinations'
- (28f) utitshala ubiza abafundi
 utitshala u-biz-a abafundi
 teacher AGRS-call-PRES students
 abangabhali viwo
 a-ba-nga-bhal-i viwo
 AFF(RC)-AGRS-NEG-write-NEG examination
 'the teacher is calling the students who are not writing (any) examination'

The relative mood can appear in all the various tenses discussed in the subsection on tense inflection.

A relative mood clause can also occur after certain conjunctives (as in (29a)), often as an alternative to the participial mood clause due to the phenomenon that the distinction between the relative mood and participial mood is vacuous in some Southern Bantu languages such as Xitsonga.

- (29a) abafundi bathula
 abafundi ba-thul-a
 students AGRS-be.quiet-PRES
 xa bafundayo
 xa ba-fund-a-yo
 when AGRS-learn-PRES-AFF(RC)
 'the students are quiet when they study'
- (29b) kuseloko abafundi bafundayo
 kuseloko abafundi ba-fund-a-yo
 since students AGRS-learn-PRES-AFF(RC)
 kakuhle baphumelela uviwo
 kakuhle ba-phumelel-a uviwo
 well AGRS-pass-PRES examination
 'since the students study well they pass the examination'

Subjunctive Mood The subjunctive mood is associated with a range of semantic contexts. It can appear in clauses denoting successive actions, necessity and obligation, purpose, wish, and prohibition, as shown in the following examples, in which these meanings are often determined by the semantic features of the verb by which it is subcategorized.

The subjunctive mood is clearly identifiable by overt morphology, specifically the verbal suffix *-e* and the subject agreement prefix *a-* for class 1 nouns. Its morphology is invariable, and it does not exhibit tense distinctions.

- Successive actions.

- (30) abantwana bavuka kusasa bahlambe
 abantwana ba-vuk-e ba-hlamb-e
 children AGRS-wake-up early
 ubuso batye
 ubuso ba-ty-e
 AGRS-wash-AFF(SUBJ) AGRS-eat-AFF(SUBJ)
 babulise abazali
 ba bulis-e abazali
 AGRS-greet-AFF(SUBJ) parents
 baye esikolweni
 ba-y-e isikolo-ini
 AGRS-go-AFF(SUBJ) LOC school-LOC
 'the children wake up early, wash (their) faces, eat, greet (their) parents, and go to school'

- Necessity, obligation.

- (31a) utitshala uyalela ukuba
 utitshala u-yalel-a ukuba

teacher AGRS-instruct-PRES COMP
 abantwana bafunde iincwadi
 abantwana ba-fund-e incwadi
 children AGRS-read-AFF(SUBJ) book
 'the teacher instructs the children to read
 a book'

- (31b) kufuneka ukuba abafundi
 ku-funeka ukuba abafundi
 EXIST-be.needed COMP students
 babhale uviwo
 ba-bhal-e uviwo
 AGRS-write-AFF(SUBJ) examination
 'it is necessary that the students write the
 examination'

• Request, wish, desire.

- (32a) utitshala ucela ukuba
 utitshala u-cel-a ukuba
 teacher AGRS-request-PRES COMP
 abantwana bafunde incwadi
 abantwana ba-fund-e incwadi
 children AGRS-read-AFF(SUBJ) book
 'the teacher requests the children to read
 a/the book'
- (32b) abazali banqwenela ukuba
 abazali ba-nqwenel-a ukuba
 parents AGRS-wish-PRES COMP
 abafundi baphumelele uviwo
 abafundi ba-phumelel-e uviwo
 students AGRS-pass-AFF(SUBJ) examination
 'the parents wish that the students pass the
 examination'

• Purpose.

- (33a) abafundi bafunda kakhulu ukuze
 abafundi ba-fund-a kakhulu ukuze
 students AGRS-learn-PRES much COMP
 baphumelele uviwo
 ba-phumelel-e uviwo
 AGRS-pass-AFF(SUBJ) examination
 'the students study hard so that they pass the
 examination'
- (33b) abafundi bafundela ukuba
 abafundi ba-fund-el-a ukuba
 students AGRS-learn-APPLIC-PRES COMP
 baphumelele uviwo
 ba-phumelel-e uviwo
 AGRS-pass-AFF(SUBJ) examination
 'the students study so that they pass the
 examination'
- (33c) abafana basebenza evenkileni
 abafana ba-sebenz-a e-ivenkile-ini
 young.men AGRS-work-PRES LOC-shop-LOC

ukuze bafumane imali
 ukuze ba-fuman-e imali
 COMP AGRS-get-AFF(SUBJ) money
 'the young men work in the shop so that they
 get money'

- Questions expressing potential necessity or obligation. Subjunctive mood interrogatives are allowed only with first-person subject pronominals.

- (34a) ndincede aba bantu?
 ndi-nced-e aba bantu
 AGRS(1.SING)-help-AFF(SUBJ) DEM people
 'must I help these people?'

- (34b) singene endlwini?
 si-ngen-e e-indlu-ini
 AGRS(1.PL)-enter-AFF(SUBJ) LOC-house-LOC
 'must we enter into the house?'

- Prohibition. A subjunctive mood clause that denotes a prohibition must have a subject pronominal in the second person and must be in the negative.

- (35a) ungayibeki incwadi apha!
 u-nga-yi-bek-i incwadi apha
 AGRS (2.SING)-NEG- book here
 AGRO-put-NEG
 'don't put the book here!'
- (35b) ningalibali ukuthenga isonka!
 ni-nga-libal-i uku-thenga isonka
 AGRS (2.PL)-neg- INF-buy bread
 forget-NEG
 'don't forget to buy bread!'

• Exhortation.

- (36a) usale kamnandi
 u-sal-e kamnandi
 AGRS(1.SING)-stay(behind)- nicely
 AFF(SUBJ)
 'you stay (behind) nicely'
- (36b) uhambe kakuhle
 u-hamb-e kakuhle
 AGRS(1.SING)-travel-AFF(SUBJ) well
 'you stay (behind) well'
- (36c) nilale kamnand
 ini-lal-e kamnandi
 AGRS(1.PL) nicely
 'you must sleep nicely'

- The subjunctive mood in the complement clause of deficient verbs.

- (37a) umfundi uphinda abhale uviwo
 umfundi u-phinda a-bhal-e uviwo
 student AGRS-do. AGRS-write- examination
 again AFF(SUBJ)
 'the student again writes the examination'

- (37b) umfundi ukhawuleza
 umfundi u-khawuleza
 student AGRS-do.quickly
 abhale uviwo
 a-bhal-e uviwo
 AGRS-write-AFF(SUBJ) examination
 'the student quickly writes the examination'

Consecutive Mood The consecutive (CONS) mood occurs in clauses that denote successive actions or states, in which the first verb is in the past tense. It is an invariable form that cannot have tense distinctions.

- (38) umntwana uvuke wahlamba
 umntwana u-vuk-e u-a-hlamba
 child AGRS-wake-PERF AGRS-AFF(CONS)-wash
 ubuso watya wabulisa
 ubuso u-a-tya u-a-bulisa
 face AGRS-AFF(CONS)-eat AGRS-AFF(CONS)-greet
 abazali waya esikolweni
 abazali u-a-ya e-isikolo-ini
 parents AGRS-AFF(CONS)-go LOC-school-LOC
 'the child woke up, washed (his/her) face, ate, greeted
 (his/her) parents, and went to school'

A consecutive mood clause may occur as complement of certain deficient verbs, in which the deficient verb itself is normally in the past tense.

- (39a) umfundi uphinde wabhala uviwo
 umfundi u-phind-e wa-bhala uviwo
 student AGRS-do. AGRS
 again- (CONS)-write
 PERF
 'the student again wrote the examination'

- (39b) umfundi ukhawuleze wabhala uviwo
 umfundi u-khawulez-e wa-bhala uviwo
 student AGRS-do. AGRS
 quickly- (CONS)-
 PERF write
 'the student quickly wrote the examination'

Imperative Mood The imperative is used for commands and instructions. If the command or instruction is directed to more than one person, the verb takes the suffix *-ni*.

- (40a) funda incwadi!
 read book
 'you (SING) read the book!'
 (40b) fundani incwadi!
 funda-ni incwadi
 read-PL book
 'you (PL) read the book!'

Hortative Mood The hortative (HORT) mood is used in clauses that express polite direct requests, in which

instance the deficient verb *-kha* is used. The hortative can also be used to express indirect requests or instructions.

- (41a) khawuphendule imibuzo
 kha-wu-phendul-e imibuzo
 let-AGRS(1.SING)- answer-AFF(HORT)
 'please answer the questions'
 (41b) khanifunde iincwadi
 kha-ni-fund-e iincwadi
 let-AGRS(1.PL)-read-AFF(HORT) books
 'please read the books'
 (41c) abafundi mabaphendule imibuzo
 abafundi ma-ba-phendul-e imibuzo
 students AFF(HORT)-AGRS- questions
 answer-AFF(HORT)
 'the students must answer the questions'
 (41d) umntwana makafunde iincwadi
 umntwana ma-ka-fund-e iincwadi
 child AFF(HORT)-AGRS- books
 read-AFF(HORT)
 'the child must read the book'
 (41e) abafundi mabangayiphenduli imibuzo
 abafundi ma-ba-nga-yi- imibuzo
 phendul-i
 students AFF(HORT)-AGRS- questions
 NEG-AGRO-answer-
 AFF(HORT)
 'the students must not answer the questions'
- The hortative does not exemplify any tense distinctions.
- Temporal Mood** The temporal (TEMP) mood occurs as a subordinate clause that denotes an activity that takes place (partly) simultaneously with the activity or state denoted by the main clause. It contains the invariable verbal prefix *-aku*. The logical subject argument usually appears in the postverbal position.
- (42a) bakufunda abafundi
 ba-aku-funda abafundi
 AGRS-AFF(TEMP)-study students
 baxoxa incwadi
 ba-xox-a incwadi
 AGRS-discuss-PRES books
 'when the students study, they discuss the books'
 (42b) sakufika ekhaya
 si-aku-fika e-ikhaya
 LOC-home AGRS-PRES-rest-PRES
 siyaphumla
 si-ya-phuml-a
 AGRS-AFF(TEMP)-arrive
 'when we arrive at home, we rest'

- (42c) bakungasebenzi abafundi badiniwe
 ba-aku-nga abafundi ba-diniwe
 -sebenz-i
 AGRS-AFF(TEMP) students AGRS-tired
 -NEG-work-NEG
 'when the students do not work, they are tired'

Infinitive Mood The infinitive mood clause is regularly subcategorized by specific (cognition) verbs, as in (43a) and (43b). It may occur in NP argument positions in a nominalized grammatical function, as in (43c) and (43d). Some verbs can allow a purposive infinitival complement only if they have an applicative suffix, as in (43e) and (43f) below (Visser, 1989).

- (43a) abafana bayakwazi ukunceda abazali
 abafana ba-ya-ku-azi uku-nceda abazali
 young.men AGRS-PRES- INF-help parents
 AGRO-know
 'the young men know (how) to help (their) parents'
- (43b) umfundi uyaqonda ukuphendula imibuzo
 umfundi u-ya-qonda uku-phendula imibuzo
 student AGRS-PRES- INF-answer questions
 understand
 'the student understands to answer questions'
- (43c) ukubhala kwabafundi kulungile
 uku-bhala kwa-abafundi ku-lungile
 INF-write GEN-students AGRS(EXIST)-good
 'the writing of the students is good'
- (43d) ukungafundi kwabafundi
 uku-nga-fund-i kwa-abafundi
 INF-NEG-learn-NEG GEN-students
 kuyamangalisa
 ku-ya-mangalisa
 AGRS(EXIST)-PRES-amaze
 'the nonlearning of the students amazes (people)'
- (43e) abafana basebenzela ukufumana imali
 abafana ba-sebenz-el-a uku-fumana imali
 young.men AGRS-work- INF-get money
 APPLIC-PRES
 'the young men work to get money'
- (43f) abafundi bafundela ukuphumelela uvio
 abafundi ba-fund-el-a uku-phumelela uvio
 students AGRS-learn- INF-pass examination
 APPLIC-PRES
 'the students study to pass the examination'

Ideophones IsiXhosa is characteristic of the Bantu languages in that it is rich in ideophones. Ideophones can function as predicates, adverbs, or interjections and are often onomatopoeic. They denote the manner or sound of an activity or the color of an object. The ideophone (IDEO) that forms part of a predicate has

inherent lexical properties of transitivity. In isiXhosa, the verb *-thi*, which co-occurs with the ideophone to form a predicate, serves as the host element for inflection, but it can be omitted in certain instances (Du Plessis, 1978, Du Plessis and Visser, 1998).

Intransitive Ideophones

- (44a) lo mntwana uhleli uthe
 lo mntwana u-hleli u-th-e
 DEM child AGRS-sit AGRS-do-PERF
 qwa
 qwa
 ideo (upright)
 'this child sat upright'
- (44b) abantu bathi nqa
 abantu ba-th-i nqa
 people AGRS-do-PRES ideo (surprise)
 ngale nto
 nga-le nto
 about-this thing
 'the people are surprised by this thing'
- (44c) ixhego lithe chu
 ixhego li-th-e chu
 old.man AGRS-do-PERF ideo (go.slowly)
 waya endlini
 wa-ya e-indlu-ini
 AGRS(CONS)-go LOC-house-LOC
 'the old man walked slowly and went to the house'
- (44d) abafazi bathe xha
 abafazi ba-th-e xha
 women AGRS-do-PERF ideo (wait)
 'the woman waited'
- (44e) le ndoda ithe xhwenene
 le ndoda i-th-e xhwenene
 this man AGRS-do-PERF ideo (suddenly.stop)
 'this man stopped suddenly'

The ideophones in (44a)–(44e) also illustrate the various click sounds, the ingressive sounds borrowed by isiXhosa from the Khoisan languages. The consonant *q* represents the palatal ingressive click sound [•], the consonant *c* represents the dental ingressive click sound [•], and the consonant *x* or (*xh*) represents the alveolateral ingressive click sound [•].

Transitive Ideophones

- (45a) lo mfana wathi
 lo mfana wa-th-i
 DEM young.man AGRS-do-PRES
 rhuthu intonga yakhe
 rhuth intonga ya-khe
 ideo (take.out) stick GEN-his
 'this young man took out his stick'
- (45b) lo mfana uthe
 lo mfana u-th-e
 DEM young.man AGRS-do-PERF

| | | | |
|-------|---------------------------------|-------------|---------------|
| | qhiwu | indebe | |
| | qhiwu | indebe | |
| | ideo (hold.high) | cup | |
| | 'this man held the cup high up' | | |
| (45c) | bamthe | hlasi | ngengalo |
| | ba-m-the-e | hlasi | nga-ingalo |
| | AGRS-AGRO-do-PERF | ideo (grab) | by-arm |
| | 'they grabbed him on the arm' | | |
| (45d) | bamthe | nqaku | lo mfana |
| | ba-m-th-e | nqaku | lo mfana |
| | AGRS-AGRO | ideo (grab) | DEM young.man |
| | -do-PERF | | |
| | 'they grabbed this young man' | | |

Bibliography

- Doke C M (1943). *Communication no. 12: Outline grammar of Bantu*. Grahamstown, South Africa: Rhodes University.
- Doke C M (1954). *The southern Bantu languages*. London: Oxford University Press for the International African Institute.
- Du Plessis J A (1978). *IsiXhosa* 4. Parow: Oudiovista.
- Du Plessis J A (1980a). 'Sentential infinitives and nominal infinitives.' *South African Journal of African Languages* 2, 50–86.
- Du Plessis J A (1980b). *Stellenbosch Studies in African Languages 2: Transitivity in Sesotho and Xhosa*. Stellenbosch, South Africa: Stellenbosch University.
- Du Plessis J A (1983). 'The quantifier *onke* in Xhosa.' *South African Journal of African Languages* 3, 79–114.
- Du Plessis J A (1986). 'Present tense in Xhosa: what does it mean?' *South African Journal of African Languages* 6, 70–73.
- Du Plessis J A (1989). 'Distribution of the complementiser *ukuba* in the Xhosa sentence.' *South African Journal of African Languages* 9, 43–51.
- Du Plessis J A (1997). *Stellenbosch Communications in African Languages 6: Imofoloji yeelwimi zaseAfrika. [Morphology of the African languages]*. Stellenbosch, South Africa: Stellenbosch University.
- Du Plessis J A & Visser M W (1992a). 'Coordination and the subjunctive in Xhosa.' *South African Journal of African Languages* 13, 74–81.
- Du Plessis J A & Visser M W (1992b). *Xhosa syntax*. Pretoria, South Africa: Via Afrika.
- Du Plessis J A & Visser M W (1998). *Stellenbosch Communications in African Languages 7: Isintaksi yesiXhosa [Xhosa syntax]*. Stellenbosch, South Africa: Stellenbosch University.
- Gowlet D (2003). 'Zone S.' In Nurse D & Philippson G (eds.) *The Bantu languages*. London: Routledge. 609–638.
- Greenberg J (1963). *Indiana University Research Center in Anthropology, Folklore and Linguistics publication 25: The languages of Africa*. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore and Linguistics.
- Guthrie M (1967–1971). *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages* (4 vols). Farnborough, UK: Gregg Press.
- Guthrie M (1971). *Comparative Bantu* (vol. 2). Farnborough, UK: Gregg Press.
- Heine B & Nurse D (eds.) (2000a). *African languages: an introduction*. Cambridge, UK: Cambridge University Press.
- Heine B & Nurse D (2000b). 'Introduction.' In Heine B & Nurse D (eds.) *African languages: an introduction*. Cambridge, UK: Cambridge University Press. 1–10.
- Louw J A (1963). *Handboek van Xhosa*. Johannesburg, South Africa: Bona Publishers.
- Nurse D & Philippson G (eds.) (2003a). *The Bantu languages*. London: Routledge.
- Nurse D & Philippson G (2003b). 'Introduction' In Nurse D & Philippson G (eds.) *The Bantu languages*. London: Routledge. 1–12.
- Piron P (1998). 'Internal classification of the Bantoid language group, with special focus on the relations between Bantu, Southern Bantoid and Northern Bantoid languages.' In Maddieson I & Hinnebusch T (eds.) *Language history and linguistic description in Africa*. Trenton, NJ: African World Press. 65–74.
- Poulos G & Msimang C T (1998). *A linguistic analysis of Zulu*. Pretoria, South Africa: Via Afrika.
- Satyo S C (1986). 'Topics in Xhosa verbal extensions.' Ph.D. diss., University of South Africa.
- Visser M W (1984). 'Aspects of empty categories in Xhosa within the theory of Government and Binding.' *South African Journal of African Languages* 5, 24–33.
- Visser M W (1986). 'Cliticization and case theory in Xhosa.' *South African Journal of African Languages* 6, 129–137.
- Visser M W (1989). 'The syntax of the infinitive in Xhosa.' *South African Journal of African Languages* 9, 154–185.
- Visser M W (1997). 'The thematic structure of event nominals in Xhosa.' *South African Journal of African Languages* 17, 65–74.
- Visser M W (2002). 'The category DP in Xhosa and northern Sotho.' *South African Journal of African Languages* 22, 280–293.
- Visser M W (2004). 'Theory-based features of task-based course design: IsiXhosa for specific purposes in local government.' *Alternation* 11(2), 216–246.
- Visser M W (2005). 'Genre analysis and task-based course design for isiXhosa second language teaching in local government contexts.' *Per Linguam* 20(1), 36–37.
- Webb V (2001). *Language in South Africa: the role of language in national transformation, reconstruction and development*. Amsterdam: John Benjamins.
- Welmers W E (1973). *African language structures*. Berkeley, CA: University of California Press.
- Williamson K & Blench R (2000). 'Niger-Congo.' In Heine B & Nurse D (eds.) *African languages: an introduction*. Cambridge, UK: Cambridge University Press. 11–42.

Y

Yakut

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Yakut (*saxa tıla*) belongs to the Northeastern or Siberian branch of Turkic languages. It has about 380 000 native speakers living in northeastern Siberia, mainly in the Yakut Autonomous Republic (*Saxa Avtonomnay Respublikata*) within the Russian Federation. The republic, whose capital is Yakutsk, has around one million inhabitants, of whom one-third are Yakuts.

The Yakut language occupies the easternmost and, together with Dolgan, the northernmost Turkic-speaking area. The huge Yakut territory has its center in the lowlands on the middle and lower reaches of Lena and its tributaries Aldan and Vilyuy; most Yakuts live in this region. In the northwest, the Yakut territory extends up to the Arctic Ocean, comprising the Khatanga river system. In the extreme northwest, speakers of Yakut live on the Taimyr peninsula, particularly on the southern slopes of the Byrranga mountain range. In the northeast, the territory extends to the lowlands of the Yana and Indigirka river systems, up to the New Siberian Islands, and even beyond the Kolyma river. Small groups of Yakut speakers live outside the republic, e.g., in the Magadan, Irkutsk, Chita, Amur, and Khabarovsk areas.

In spite of the strong dominance of Russian, which is the language of higher education, Yakut has a relatively strong status in the republic, also being used as a second language by many speakers of Evenki, Even, and Yukagir.

Origin and History

After their emigration to northeastern Siberia, the ancestors of the Yakuts lost their contact with other Turkic-speaking groups. Since their language has been geographically isolated from other Turkic varieties for many centuries, it exhibits features that sharply distinguish it from them and makes it unintelligible

to speakers of other Turkic languages. Numerous archaic features show that the contact with the rest of the Turkic world was lost very early. On the other hand, a great number of deviations are the results of innovative developments.

The ancestors of the Yakuts seem to have belonged to the 'tree Kurikan tribes' (*üč qurıqan*) mentioned in the East Old Turkic stone inscriptions found in the Orkhon river valley. It is obvious that they lived for a relatively long time in the area surrounding Lake Baikal before they migrated northward. This is also indicated by the Yakut word *bayaɣal* 'sea.' Various Turkic-speaking groups have settled in the Baikal region, also the ancestors of the Tuvans and old Uyghur groups. The Yakut language itself contains indications of an early habitat in the south, e.g., names of months that do not fit the climate of Yakutia and words for animals such as *tebien* 'camel.' Yakut oral traditions also tell us about a migration from the south to the north. Early Yakut tribes left their southern habitat, probably pushed by Buryat groups, and migrated northward along the Lena river. This exodus did not occur before the 13th century, since the memory of Chinggis and the Mongol campaigns is still alive in Yakut traditions.

The ancestors of the Yakuts had been subject to a certain Mongolian admixture prior to the migration. When proceeding northward along the Lena river, the Turkic-speaking immigrants mixed with and absorbed indigenous Evenkis, Evens, and Yukagirs. At the same time, they also pushed local Tungusic-speaking groups northwestward and northeastward. Yukagirs and Paleoasiatic groups were forced out to still more peripheral regions. For centuries, however, the Yakuts lived south of their present-day territory. It was only under the pressure of the Russian expansion in Siberia that they migrated to more arctic regions. Yakutia was incorporated in the Russian Empire in the 1620s.

While the Yakuts have preserved many features of the southern culture of cattle and horse breeders, they have also taken over elements of northern nomadism from their new neighbors, traditionally reindeer herders and hunter-gatherers. In spite of Christianization and Russification, their ethnic structure has remained relatively intact.

Related Languages and Language Contacts

Yakut is most closely related to its geographically nearest Turkic neighbors, Tuvan and Khakas of southern Siberia. The old Yakut self-designation *Ura:nxay* points to early connections with the territory of Tuva, which has also been referred to as Uryankhay. Some scholars have assumed that Yakut was originally a Kipchak Turkic language (see **Turkic Languages**).

Yakut has been in long and close contact with other languages. It shows strong traces of Mongolic influence. The period in which the ancestors of the Yakut settled on the shore of Lake Baikal led to close interaction with Buryat. An early impact on Yakut may also have been exerted by Yeniseian, a formerly widespread Paleoasiatic language. After the emigration to northern Siberia, the Turkic language of the Yakuts underwent strong substrate influence from Tungusic dialects. The next neighbors of Yakut are the North Tungusic languages Evenki and Even (Lamut), both of which appear to have Paleoasiatic substrates. The speakers of Evenki live in the northern and northwestern parts of Yakutia, whereas the Evens live in the northeastern parts, in particular in the basins of the rivers Indigirka, Yana, and Kolyma. The contacts with the isolated language Yukagir have also been important. The complex problems of language contact and language shift in the area are still unsolved.

The Written Language

No old Yakut literary documents are known. According to a tradition in Yakut folklore, however, the Yakuts once possessed written documents, which they lost on their way to the north. There is a rich Yakut oral literature comprising legends, epics, songs, etc. A modern literature began to develop at the beginning of the 20th century. A Cyrillic alphabet was created for Yakut by the German scholar Otto Böhtlingk in the mid-19th century. A new script, based on the International Phonetic Alphabet and designed by the Yakut linguist S. A. Novgorodov, was introduced in 1922. It was later replaced by a new Roman-based script, which was in use until a Cyrillic alphabet was introduced in 1939. The orthographic rules of the modern Yakut language have often been changed. They have, however, basically followed phonetic principles, mirroring the actual pronunciation with its numerous assimilations.

Distinctive Features

Yakut exhibits many linguistic features typical of the Turkic family (see **Turkic Languages**). It has, for

example, a suffixing morphology, sound harmony, and a head-final constituent order. In the following, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation. Hyphens are used here to indicate morpheme boundaries.

Phonology

Yakut holds an exceptional position among the Turkic languages because of certain phonetic developments. Similar phenomena are sometimes found in contact languages such as Buryat and Evenki.

Like Turkmen and Khalaj in the southwestern part of the Turkic-speaking world, Yakut has preserved Proto-Turkic long vowels, e.g., *a:t* 'name' and *ü:t* 'milk.' Yakut has eight short vowels and eight long vowels including four diphthongs. The nonhigh long vowels are realized as diphthongs, e.g., *küöl* 'lake' < *kö:l*. Yakut *t* corresponds to the East Old Turkic intervocalic and word-final dental *ð*, e.g., *atax* 'leg,' *tot-* 'to become satiated.' Initial *s-* corresponds to *y-* in most other Turkic languages, e.g., *suol* 'way,' *sit-* 'to lie' (Turkish *yol*, *yat-*). The consonants *z*, *š*, and *č* have developed into *s* in Yakut, e.g., *seri:* 'army' < *čerig*. Initial *s-* has been deleted, e.g., *u:* 'water,' *ös* 'word'; cf. Turkish *su*, *söz*. Intervocalic *-s-*, however, has developed into *-h-*, e.g., *kub-a* [duck-POSS.3.SG] 'his/her duck' (of *kus* 'duck'), *uhun* 'long'; cf. Turkish *kuş* 'bird,' *uzun* 'long.'

Yakut applies, like other Turkic languages, a front-back sound harmony, according to which native words contain either front or back sounds. The rounded-unrounded harmony is also well developed. The vowels *o* and *ö* may occur as suffix vowels, e.g., *kötör-ö* [bird-POSS.3.SG] 'his/her bird' (*kötör* 'bird').

Due to sound changes, progressive and regressive consonant assimilations, unstable vowels, etc., Yakut word forms often deviate from the typical Turkic agglutinative structure, e.g., *at* 'horse' vs. *ap-pit* [horse-POSS.1.PL] 'our horse,' *tayis-* 'to go out' vs. *taxs-ar* [go out-PRES.3.SG] 'goes out,' *kis:* 'daughter' vs. *kih-im* [daughter-POSS.1.SG] 'my daughter.' Some pronouns have special oblique stems, e.g., *mi:gi-* vs. *min* 'I,' *man-* vs. *bu* 'this.' The third-person imperative form consists of the verbal stem, e.g., *as* 'open,' whereas the corresponding negative form exhibits a vowel element in front of the negation marker, e.g., *ah-i-ma* [open-i-NEG.IMP] 'don't open'; cf. Turkish *aç* [open.IMP], *açma* [open-NEG.IMP].

Grammar

Yakut displays some unique grammatical features, innovations partly due to Mongolic and/or Tungusic

influence. Striking features in the case system are the lack of a genitive and the fusion of dative and locative. The nominative is used instead of a genitive in constructions such as *kihi bihaγ-a* [man knife-POSS] 'the man's knife.' The old locative-ablative has lost its spatial functions, assuming a partitive function with imperatives and necessitives, e.g., *u:-ta ayal* [water-PART bring-IMP.2.SG] 'bring [some] water.' Its locative function has been taken over by the dative-locative in -GA, which expresses both location and goal, e.g., *guorak-ka* [town-DAT.LOC] 'in/to the town.' Special case suffixes occur with possessive markers. For instance, while *ak-ka* [horse-DAT.LOC] 'to the horse' is the dative-locative form of *at* 'horse,' *ap-p-ar* [horse-POSS.1.SG-DAT.LOC] 'to my horse' is the corresponding form of *at-im* 'my horse.' New cases have emerged in Yakut, an instrumental, a comparative, a comitative, and an adverbial case. An example of the latter is *kihi-li* [human being-ADV] 'in a human way' (*kihi* 'human being').

The Yakut yes-no question marker is *duo* or *du:*, whereas almost all other Turkic languages use markers of the type *ml*. An enclitic question particle *-iy* is added to interrogative pronouns and adverbs, e.g., *bu kimiy?* [this who-INTERROG] 'who is this?' Possession may be expressed by means of the adjective suffix *-la:x*, e.g., *Min ĵie-le:x-pin* [I house-PROVIDED.WITH-1P.SG.] 'I have a house.' The adjective *suox* 'nonexisting' (cf. Turkish *yok*) is used instead of the common Turkic privative suffix *-siz* 'without,' e.g., *u:-ta suox* [water-POSS nonexisting] 'without water'; cf. Turkish *su-suz* [water-PRIV]. Adjectives may be negated with a third-person possessive suffix plus *suox*, e.g., *kuhayaana suox* [bad-POSS.3.SG nonexisting] 'not bad'. The cardinals numbers from 11 to 19 are formed with *uon* 'ten' plus a digit, e.g., *uon tüört* [ten four] 'fourteen.' The tens from 40 to 90 are formed with a digit plus *uon* 'ten,' e.g., *tüört uon* [four ten] 'forty'; cf. Turkish *kirk*.

An archaic feature is the retention of the verbal suffix *-BIt*, which is otherwise only found in the southwestern branch of Turkic, e.g., *kel-bit* [come-PART] 'having come'; cf. Turkish *gel-miş* [come-PART]. As a finite form it has evidential (indirective) meaning, e.g., *kel-bit* [come-EV.PAST.3.SG] 'has evidently come'; cf. Turkish *gel-miş* [come-EV.PAST.3.SG] 'has evidently come.'

Lexicon

The basic lexicon of Yakut is of Turkic origin. Most words of foreign origin are Mongolic loans. There is an old Buryat layer from the early period of settlement on the shore of Lake Baikal. Even the

pronoun *beye* 'self' has been copied from Mongolic. Loanwords from Tungusic often belong to the domain of husbandry and everyday life. A large portion of the Yakut lexicon is of unknown origin, probably due to contact with Paleoasiatic languages. The Russian impact on the lexicon has been considerable. Loanwords are in general assimilated to the Yakut word structure, e.g., *silaba:r* 'samovar,' *biragra:mma* 'programme.'

Dialects

The differences between the Yakut dialects are comparatively small. There is a central group consisting of the Aldan, eastern and western Lena dialects, a northeastern group, influenced by Even, and a northwestern group, influenced by Evenki.

Another dialect is Dolgan, spoken by about 6500 persons, mainly on the Taimyr peninsula. It differs from the northwestern dialects and has its origin in Tungusic groups who shifted to Yakut at an early stage. They left their settlements on the Vilyuy at the end of the 16th century or later, migrated northward and absorbed parts of the population of the Taimyr peninsula, primarily Nganasans, i.e., speakers of Tavgi Samoyedic, and also further groups. Dolgan thus has both an Evenki and a Nganasan substrate. It still functions as the lingua franca of Taimyr. The present-day Dolgans (self-designation *haka*, corresponding to *saxa*) distinguish themselves from Yakuts and consider their variety a language in its own right. Dolgan differs somewhat from other Yakut varieties in lexical respect, and it also displays a few differences in terms of phonology and grammar. An archaic feature is the absence of the change of initial and final *q* to *x*, e.g., *katun* 'woman' (Yakut *xotun*), *kol* 'shoulder' (Yakut *xol* 'arm'), *atak* 'foot' (Yakut *atax*), *huok* 'nonexisting' (Yakut *suox*). An innovative feature is the development of secondary *s-* into *h-*, e.g., *haka* 'Yakut' (Yakut *saxa*), *hil* 'year' (Yakut *sil*), *heri*: 'war' (Yakut *seri*:).

Bibliography

- Böhtlingk O (1851). *Über die Sprache der Jakuten*. St. Petersburg.
- Buder A (1989). *Aspekto-temporale Kategorien im Jakutischen*. Turcologica 9. Wiesbaden: Harrassowitz.
- Kałużyński S (1961). *Mongolische Elemente in der jakutischen Sprache*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Krueger J R (1962). *Uralic and Altaic Studies 21: Yakut manual*. Indiana University Publications. The Hague: Mouton.
- Pekarskij È K (1907–1930). *Slovar' jakutskogo jazyka*. St Petersburg/Leningrad: Akademija Nauk.

Poppe N (1959). 'Das Jakutische.' In Deny J *et al.* (eds.) *Philologiae turcicae fundamenta* 1. Wiesbaden: Steiner. 671–684.

Stachowski M (1993). *Dolganischer Wortschatz*. Kraków: Uniwersytet Jagielloński.

Stachowski M & Menz A (1998). 'Yakut.' In Johanson L & Csató É Á (eds.) *The Turkic languages*. London/New York: Routledge. 417–433.

Ubrjatova E I (ed.) (1982). *Grammatika sovremennogo jakutskogo literaturnogo jazyka*. Moscow: Nauka.

Yanito

D Levey, Universidad de Cádiz, Cádiz, Spain

© 2006 Elsevier Ltd. All rights reserved.

Yanito (or Llanito) is the name commonly used to refer to the people of Gibraltar as well as their local vernacular. Although various theories exist, it seems likely that it has its etymological origins either in the English name 'Johnny' or alternatively, reflecting the traditional Genoese presence in the British colony, it may be derived from 'Gianni,' the diminutive of the Italian boys' name 'Giovanni.'

Yanito is not an autonomous language as such, and it is seldom found in written form. It is fundamentally a spoken Spanish-dominant variant, which incorporates English lexical and syntactic constituents as well as some unique local lexical items. Although the Spanish/English content ratio may vary from speaker to speaker, most Gibraltarians will, consciously or unconsciously, alternate between English and Andalusian Spanish in everyday situations.

Code-switching may take place inter-sententially or intra-sententially.

Yanito

What? Pero ... I told you, no? No puedo ir shopping porque I have to work late. Sorry, no puedo hablar ahora. Anyway, te llamo esta noche OK?

English

What? But ... I told you, didn't I? I can't go shopping because I have to work late. Sorry, I can't speak now. Anyway I'll phone you tonight, OK?

Although L1 interference and unnatural direct translations may sometimes be present resulting in what is popularly known as 'Spanglish,' the syntactic rules of both languages tend to be respected.

Individual English lexical items, particularly nouns, are commonly introduced in otherwise Spanish utterances. This is often because no direct equivalent exists or its cultural or social nuance cannot be easily or succinctly conveyed. Although British English pronunciation norms tend to be followed, some older borrowings and derivations have been Hispanicized, usually reflecting the local Andalusian pronunciation. Interestingly, several of these words have also found

their way across the border and are used in the neighboring Spanish towns of La Linea and San Roque.

- *chinga* = chewing gum
- *liqueribá* = liquorice bar
- *mebli* = marbles
- *el tishella tisha* = teacher

Other English borrowings have taken on a different meaning within the local context. *Pish-pine* (from the English 'pitch pine'), for example, is used locally as an adjective or an adverb to mean 'perfect.'

Al final todo salió pish-pine.

In the end everything turned out just fine.

While Spanish and English form the basis of the local lexicon, several borrowings from other languages are also present, reflecting the multicultural makeup of the British colony. These come mainly from Italian (e.g., *pompa* = pump), Arabic (e.g., *flush* = money), and Hebrew (e.g., *ha ham* = boss).

Although Yanito does not hold prestige status, it is not overtly stigmatized either. It is regarded with a certain degree of affection and used by many Gibraltarians as an expression of local identity. However, although the use of Yanito is widespread and considered by many to be a defining characteristic of the local speech community, Gibraltar can not be described as a diglossic speech community. Both English and Andalusian Spanish are very much alive, and speakers may adopt either of the three language forms depending on context, domain, and the interlocutor.

Bibliography

Ballantine S (2000). 'English and Spanish in Gibraltar: development and characteristics of two languages in a bilingual community.' *Gibraltar Heritage Journal* 7, 115–124.

Cavilla Manuel (1990). *Diccionario Yanito*. Gibraltar: MedSUN.

Fernández Martín C (2003). *Valoración de las actitudes lingüísticas en Gibraltar*. Madrid: Umi-ProQuest information on Learning.

García Martín J M (1996). *Materiales para el estudio del español de Gibraltar. Aproximación sociolingüística al*

- léxico español de los estudiantes de enseñanza secundaria.* Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- García Martín J M (1997). 'El español en Gibraltar. Panorama general.' *Demófilo. Revista de cultura tradicional de Andalucía* 22, 141–154.
- García Martín J M (2000). 'Los conceptos de bilingüismo y diglosia y la situación lingüística de Gibraltar.' *Actas del XII Congreso de la Asociación Internacional de Hispanistas, Madrid* 3, 483–489.
- Kellermann A (2001). *A new new English: language, politics and identity in Gibraltar.* Heidelberg: Herstellung.
- Kramer J (1986). *English and Spanish in Gibraltar.* Hamburg: Helmut Buske Verlag.
- Kramer J (1998). 'Die Sprache Gibraltars. Le Yanito.' In Holtus G, Metzeltin M & Schmitt C (eds.) *Lexikon der Romanistischen Linguistik. Kontakt, Migration und Kuntsprachen. Kontrastivität, Klassifikation und Typologie III.* Tübingen: Niemeyer. 310–316.
- Levey D (2004). *English pronunciation and production tendencies in Gibraltar.* Ph.D. dissertation, University of Cádiz.
- Martens J (1986). *Gibraltar and the Gibraltarians: the social construction of ethnic and gender identities in Gibraltar.* Ph.D. dissertation, University of London, School of Oriental and African Studies.
- Moyer M (1993). *Analysis of code-switching in Gibraltar.* Bellaterra: Publicacions de la Universitat Autònoma de Barcelona.
- Moyer M (1998a). 'Bilingual conversation strategies in Gibraltar.' In Auer P (ed.) *Codeswitching in conversation, language and identity.* London: Routledge. 215–234.
- Moyer M (1998b). 'Entre dos lenguas: contacto de inglés y español en Gibraltar.' In Muysken P (ed.) *Sociolingüística, lenguas en contacto, Foro Hispánico*, 13. Amsterdam/Atlanta, GA: Editions Rodopi B. V. 9–26.
- Vallejo T (2001). *The Yanito Dictionary.* Gibraltar: Panorama Publishing.

Yiddish

R A Rothstein, University of Massachusetts, Amherst, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Yiddish, a Germanic language spoken by the Jews of Central and Eastern Europe (Ashkenazim) and in the Ashkenazic diaspora around the world, includes significant Semitic and Slavic components as well as its Germanic base. It is one of a number of Jewish languages created on the basis of the coteritorial non-Jewish language (cf. Judaeo-Arabic, Judaeo-Persian, Judezmo [Ladino], etc.). Of all such languages, it achieved the widest range of functions, the most highly developed network of institutions and the largest number of speakers.

Orthography

Like all Jewish languages, Yiddish is written in the Hebrew alphabet. Unlike Hebrew, which except for special purposes is normally written without vowel symbols, Yiddish has adapted certain Hebrew letters (sometimes in combination with Hebrew subscript or superscript vowel symbols) to represent vowels. Words and morphemes of Semitic (Hebrew and Aramaic) origin are for the most part spelled as they are in the source languages, while words of Germanic, Slavic, and other origins are spelled in a broadly phonetic manner. The orthography is super-dialectal, which permits its use by speakers of the standard language (the phonetics of which is based

largely on Northeastern Yiddish) as well as by speakers of dialects, which most strongly differ from the standard and one another in their realization of the vowels. In the Soviet Union, the official orthography for Yiddish 'naturalized' the Semitic component, eliminating the traditional Hebrew and Aramaic spellings in favor of the kind of phonetic representations used elsewhere for the non-Semitic component. Soviet orthography also mandated standardized representations for elements that show dialectal variation, e.g., orthographic *oyf* was spelled *af* as a preposition and *uf* as a prefix. (The standard scholarly transcription for Yiddish is the system developed by the YIVO Institute for Jewish Research [originally in pre-1939 Wilno, Poland, now in New York].)

Phonology

The phonemic inventory of standard Yiddish consists of eight vocalic segments (five oral vowels and three diphthongs) and twenty-nine consonantal segments (some of which play only a marginal role). The oral vowels are *i*, *e*, *a*, *o*, and *u*, realized phonetically as [i], [ɛ], [ɐ], [ɔ], and [u]; the diphthongs are *ey*, *ay*, and *oy*. The basic consonantal inventory contains voiced and voiceless bilabial, dental, and velar stops; bilabial, dental, and palatal nasals; voiced and voiceless labiodental, dental, and alveolar fricatives; a voiceless velar fricative and a voiced laryngeal fricative; voiceless dental and alveolar affricates; a dental and a palatal lateral; a palatal glide; and an /r/ that can

be pronounced as a uvular (most speakers) or dental trill. Voiced dental and alveolar affricates, and – regionally – palatal (or palatalized) versions of /t/, /d/, /s/, and /z/ play a somewhat marginal role in the phonological system. The resonants /l/ and /n/ can be syllabic, as can the positional variants of /n/, [m] (in word-final position after /b/ or /p/) and [ŋ] (in word-final position after /k/ or /g/).

As in Slavic, Yiddish obstruents assimilate (generally regressively) with respect to voice, but /v/ does not cause voicing in a preceding voiceless obstruent. Word-final obstruents do not (as in Slavic or German) lose voicing before pause. There is, however, evidence of such devoicing being operative in an earlier stage of the language, e.g., *hunt* ‘dog’ vs. *briv* ‘letter’ (cf. German *Hund* with final [t] and *Brief* with final [f]).

Word stress tends toward the penultimate, but words of Germanic origin are often stressed on the initial root syllable, and words of Slavic origin often retain the stress of the source language. Posttonic vowels are generally reduced.

Morphology

Three noun genders (masculine, feminine, neuter) are distinguished in the singular by agreeing forms of the definite article and attributive adjectives, as well as by pronominal reference. Northeastern Yiddish, however, like the neighboring Lithuanian language, has lost the neuter gender. There are no gender distinctions in the plural.

Nouns are pluralized by means of several endings, mostly independent of gender: *-n* or its variant *-en*, *-er*, zero (Germanic in origin); *-s* (Romance in origin); *-im*, *-es* (Hebrew in origin). The Germanic and Hebrew suffixes may be accompanied by vowel changes in the stem; the same changes take place in suffixal derivation, e.g., in diminutive formation (cf. *barg* ‘mountain,’ *berg* ‘mountains,’ *bergl* ‘hill’; *hoyz* ‘house,’ *hayzer* ‘houses,’ *hayzl* ‘little house’).

The definite article, attributive adjectives, and personal pronouns distinguish nominative, accusative, and dative case forms (with extensive syncretism) in the singular (pronouns also in the plural); personal names and a few common nouns can add *-(e)n* to indicate the accusative or dative singular and *-(e)s* to mark a possessive form. Agreeing elements have the same form for dative and possessive. Predicate adjectives occur either as a bare stem or with the indefinite article and a gender ending: *zi iz yung* ‘she is young’ vs. *zi iz a yunge* ‘she is a young one,’ parallel to Russian constructions with short- and long-form adjectives (*ona moloda* vs. *ona molodaia*). Adverbs have the same form as the stem of the corresponding adjective.

Yiddish verbs have synthetic forms for the present tense and analytic forms for the past and future tenses. The future is formed by combining the conjugated auxiliary *veln* with the infinitive; the past is formed by combining the auxiliaries *hobn* ‘have’ or *zayn* ‘be’ with the past participle. Conditional and subjunctive moods are also formed with auxiliaries: *voltn* (with the past participle) for the former and *zoln* ‘should’ (with the infinitive) for the latter. The auxiliary *flegn* ‘used to’ combines with infinitives to express iterativity in the past. A large number of periphrastic verbs combine one of several auxiliaries with an invariable element, often a Hebrew verbal form (e.g., *mekane zayn* ‘be envious’; *moyre hobn* ‘be afraid, fear’; *vey ton* ‘hurt’; *geboyryn vern* ‘be born’).

Yiddish verbs may be combined with a variety of stressed adverbial complements that are prefixed to the infinitive and participles, but follow the inflected verb as a separate word in the present tense and the imperative (e.g., *avekgeyn* ‘to go away,’ *ikh bin avekgegangen* ‘I went away’ vs. *ikh gey avek* ‘I’m going away,’ *gey avek!* ‘go away!’). Under the influence of Slavic verbal systems, some of these inherited Germanic verbal complements are used to express aspectual or Aktionsart meanings, in ways that differ both from the Germanic and Slavic systems. Yiddish also makes broad use of the so-called stem construction, which combines an auxiliary (usually *gebn* ‘give’ or *ton* ‘do’) with the indefinite article and an invariant verbal stem to create a semelfactive meaning: *gebn a kuk* ‘take a look,’ *a trakht ton* ‘give a bit of thought.’

Syntax

Yiddish is a verb-second language, with the inflected verb serving as the second clause constituent. In a complex sentence, however, an initial clause can occupy the first constituent position and the verb will therefore occupy the first position in the second clause: cf. *er vet farshteyn* ‘he will understand’ vs. *az er vet zayn elter, vet er farshteyn* ‘when he is [will be] older, he will understand.’ The verb can also occupy the first position in a clause or sentence that follows as an implied consequence of a preceding clause or sentence, e.g., *der tate iz geshtorbn, bin ikh geblibn aley n* ‘my [the] father died, [so] I was left all alone.’ Interrogative elements (the particle *tsi* that introduces yes-no questions, pronouns, adverbs) count as the first constituent in a direct question but not in an indirect question: cf. *vos hot zi geshribn?* ‘what has she written?’ vs. *ikh veys nit, vos zi hot geshribn* ‘I don’t know what she has written.’

Aside from the verb-second requirement, word order is relatively free and is available to express such discourse functions as emphasis, contrast, topic vs. comment. In order to move a subject to a more emphatic position without violating the verb-second principle, the neuter pronoun *es* (or its variants in this function *se*, *s*) serves as a dummy occupying the first constituent position, e.g., *es iz tsu mir gekumen a kuzine* ‘a cousin came to me.’

Like its Slavic counterparts (but unlike the situation in Germanic), the reflexive pronoun *zikh* serves for all persons and numbers. It also serves both as the full accusative/genitive and dative of the reflexive/reciprocal pronoun (cf. Polish *siebie/sobie*) and as the enclitic form that occurs with verbs in a variety of functions (cf. Polish *sie*). Verbs with *zikh* can express, among other things, a kind of middle voice (e.g., *vashn zikh* ‘wash/wash up/get washed’) and also an intransitive verb with an unaccusative subject (e.g., *der vinter heybt zikh on* ‘winter is beginning’). Following Slavic models, prefixed (or complemented) verbs with *zikh* express various Aktionsart meanings (e.g., *tselakhn zikh* ‘burst out laughing’; cf. Russian *rassmeiat’sia*).

Yiddish does not drop subject pronouns, although some contracted forms are used in speech (*kh* for *ikh* ‘I,’ *r* for *er* ‘he’). Second-person plural pronouns and verb forms are used in nonfamiliar address.

Lexicon

Although the basic stock of Yiddish vocabulary is Germanic in origin, there is also a large Semitic component (from Hebrew and Aramaic, known collectively in Yiddish as *loshn-koydesh* ‘the language of holiness’), which may reach as much as 15% or more depending on style and register. The significant Slavic component comes primarily from Polish, Ukrainian, and Belarusian, and there are traces of old Romance influences. Many Greek- and Latin-based internationalisms entered the language in the 19th and 20th centuries, often via one Slavic language or another. The Germanic, Semitic, Slavic, and other elements were phonologically and morphologically integrated into the Yiddish linguistic system, often being creatively reworked. The verb *balebatevn* ‘keep house; manage; bully,’ for example, contains two Hebrew roots (meaning ‘master of the house’), a Slavic suffix used to derive verbs from foreign roots (cf. Polish-*owa-*), and a Germanic infinitival ending. Calques were created both on the word level (see the above examples of verbal prefixation) and on the phrase level. A colloquial phrase meaning ‘put in jail,’ *araynzetsn in koze*, borrows the Polish slang term

for ‘jail,’ *koza*, literally, ‘goat’ and uses Germanic elements (verbal complement *arayn*, root *zets-*, infinitive ending *-n*, preposition *in*) to calque the entire Polish phrase *wsadzić do kozy*. A more elaborate version replaces the Germanic verbal root with a Semitic one, and the Polish slang term with the Aramaic-origin phrase *khad-gadye* ‘a single kid,’ giving *araynyashvenen in khad-gadye*.

Although many words from the Semitic component are related to Jewish religious life, many are not (e.g., *khaver* ‘friend; [political] comrade’; *balebos* ‘master of the house; boss’), and there is no neat correspondence between the origins of words and their sphere of application. So, for example, the verb meaning ‘say a blessing’ is *bentshn*, which is of Romance origin (ultimately related to Latin *benedicere*), while the word *got* ‘God’ is from the Germanic component (and has an affective form with a Slavic suffix, *gote-nyu*). Most kinship terms are of Germanic origin, but *zeyde* ‘grandfather’ and *bobe* ‘grandmother’ come from the Slavic component.

History

Yiddish is generally assumed to have begun to develop as a distinct linguistic variety around the year 1000 C.E. The long-dominant theory of origins (connected with the work of Max Weinreich) attributes this development to the migration of Jewish speakers of Old French and/or Old Italian, who were literate in Hebrew/Aramaic, into the Rhine Valley, where they encountered Germanic speakers. In recent years, scholars have questioned parts of this theory, suggesting Northern Italy or Bavaria as the point of initial contact, arguing that Yiddish developed as a relexification with Germanic materials of a kind of Judaeo-Slavic (Paul Wexler) or proposing that Yiddish began with contacts between Aramaic-speaking Jews from the Middle East and Germanic speakers (Dovid Katz).

As Jews moved eastward, they settled among speakers of Slavic languages (first Czech, then Polish, later Ukrainian and Belarusian). From around 1500 and until World War II, the majority of Yiddish speakers inhabited the largely Slavic-speaking lands of Central and Eastern Europe. In addition to the Yiddish-speaking religious institutions (educational, social, etc.) that functioned throughout that territory, there existed during the 1920s and 1930s in Poland and (until the mid-1930s) in the Soviet Union a wide array of educational, cultural, social, and political institutions with Yiddish as their language of instruction, publication, daily business, etc. The Nazi annihilation of European Jewry, together with

assimilation (voluntary or forced) to the dominant cultures in the Soviet Union and the overseas lands of the Eastern European Jewish diaspora, has led to a great diminution in the number of speakers of Yiddish. Yiddish is alive today among a decreasing number of elderly Jews of East European origin and in certain traditionalist (mostly Hasidic) communities in North America, England, and Israel, where it serves as the principal vernacular. It is also cultivated by an unknown number of relatively young, largely secular, Jews (and some non-Jews), who are devoted to keeping the language and culture alive.

The oldest dated Yiddish text (1272) is a sentence written in a prayer book in Worms, Germany. The first printed text is a Yiddish translation of a Hebrew prayer included in a 1526 Prague haggadah, and the first Yiddish book is a Hebrew–Yiddish Bible concordance published in Cracow in 1534 or 1535.

Dialectology

The dialect map of Yiddish is divided into Western and Eastern Yiddish, with the latter subdivided into Central (Polish), Northeastern (Lithuanian), and Southeastern (Ukrainian) Yiddish. Western Yiddish is defined roughly as Yiddish spoken west of the 1939 Polish–German border; it is the descendent of the earliest Yiddish, and even by 1939 had largely been replaced by German, although some speakers continued to use it in such areas as Alsace, Switzerland, and Slovakia. Linguistically, the Yiddish dialect continuum is divided on the basis of the development of certain proto-Yiddish vowels. In particular, the phrase ‘to buy meat’ (*koyfn fleysh* in Standard Yiddish) would be *ka:fn fla:sh* (with long vowels) in WY, *koyfn flaysh* in CY, *keyfn fleysh* in NEY and *koyfn fleysh* in SEY. Standard Yiddish (which is, strictly speaking, Standard Eastern Yiddish) is largely based on NEY as far as its vocalism is concerned, although in the case of the diphthong of ‘to buy,’ the standardizers chose the variant common to CY and SEY (*oy*) rather than the NEY variant (*ey*). The more usual choice of vowel for the standard is reflected in a phrase like ‘one day’: Standard and NEY *eyn tog*, CY *ayn tug*, SEY *eyn tug*.

Standard Yiddish, like NEY, but unlike CY and SEY, does not distinguish vowel length. It does, however, distinguish dental and alveolar fricatives and affricates, like CY and SEY, but unlike NEY. NEY also has no neuter gender, unlike the other Yiddish dialects (and the standard language). A detailed account of Yiddish dialect phenomena is presented in the multivolume *Language and culture atlas of Ashkenazic Jewry*, three volumes of which have been published as of 2004.

Bibliography

- Birnbaum S A (1979). *Yiddish: a survey and a grammar*. Toronto/Buffalo: University of Toronto Press.
- Estraikh G (1999). *Soviet Yiddish: language planning and linguistic development*. Oxford/New York: Clarendon Press/Oxford University Press.
- Estraikh G & Krutikov M (eds.) (1999). *Yiddish in the contemporary world*. Oxford: European Humanities Research Center/Oxford University.
- Herzog M I et al. (eds.) (1992). *The language and culture atlas of Ashkenazic Jewry. 1: Historical and theoretical foundations*. Tübingen: Max Niemeyer Verlag.
- Jacobs N G (2005). *Yiddish: a linguistic introduction*. New York: Cambridge University Press.
- Katz D (ed.) (1987). *Origins of the Yiddish language*. Oxford/New York: Pergamon Press.
- Katz D (ed.) (1988). *Dialects of the Yiddish language*. Oxford/New York: Pergamon Press.
- Katz D (2004). *Words on fire: the unfinished story of Yiddish*. New York: Basic Books.
- Röll W & Neuberg S (eds.) (1999). *Jiddische Philologie: Festschrift für Erika Timm*. Tübingen: Max Niemeyer Verlag.
- Weinreich U (ed.) (1954). *The field of Yiddish: studies in Yiddish language, folklore, and literature*. New York: Linguistic Circle of New York. [Subsequent volumes, various editors and publishers, 1965, 1969, 1980, 1993.]
- Weinreich M (1980). *History of the Yiddish language*. Shlomo Noble & Joshua A Fishman (trans.). Chicago/London: University of Chicago Press.
- Wexler P (ed.) (1990). *Studies in Yiddish linguistics*. Tübingen: Max Niemeyer Verlag.
- Wexler P (1993). *The Ashkenazic Jews: a Slavo-Turkic people in search of a Jewish identity*. Columbus, Ohio: Slavica Publishers.

Yoruba

K Owolabi, University of Ibadan, Ibadan, Nigeria

© 2006 Elsevier Ltd. All rights reserved.

Location and Number of Speakers

Yoruba is spoken as a first language in Nigeria in virtually all areas in the states of Èkìtì, Lagos, Ògùn, Òndó, Òşun, and Òyó, and in most of the areas in Kwara and Kogi states; Yoruba is a second language in some areas of the Delta and Edo states as well as in the non-Yoruba-speaking areas of the Kwara and Kogi states (see Figure 1). Outside Nigeria, there are Yoruba-speaking communities in the republics of Togo and Benin, where, in the southern part, Yoruba, Aja, and Fon are the three dominant indigenous languages (see Adeniran, 2004: 437). Based on the 1991 census, the number of speakers of Yoruba as a first or second language in Nigeria alone is about 19 000 000.

Genetic Relationship and History

Yoruba is a member of the Defoid language group within the Benue-Congo subgroup of the Niger-Congo family (see Williamson, 1989: 20, 26; Capo, 1989: 275–290). Although the origin of the term ‘Yoruba’ is still shrouded in mystery, some late 20th-century (historical/comparative) linguists have suggested that the dispersal center for the Yoruba people

is around the southwest area of the confluence of the Niger and Benue rivers in Nigeria (see Akinkugbe, 1978; Williamson, 1989: 270).

Earliest Written Record

The Yoruba writing system uses the Roman alphabet, augmented by letters with diacritics. The earliest written records include the vocabularies compiled by Thomas Bowdich in 1819 (including words for the numerals 1–10), by Hannah Kilham (1828), and by Wilhelm Koelle (1854). The teaching booklets of John Raban (1830–1832) and the vocabularies and grammars of Samuel Crowther (1843, 1852) contributed further to the written record. Publication of Crowther’s school primer (1849; written wholly in Yoruba) was followed by the various translation works in the Old and the New Testaments by Crowther (1950–1956) and by Thomas King (1957–1961). The first vernacular periodical, the newspaper *Ìwé Ìròhin*, was printed at Abẹokuta from 1859 to 1867 (see, in particular, Hair (1967)).

Individual Characteristics

Yoruba has a number of characteristics that appear unique to Yoruba or are not widespread among its genetic relatives.

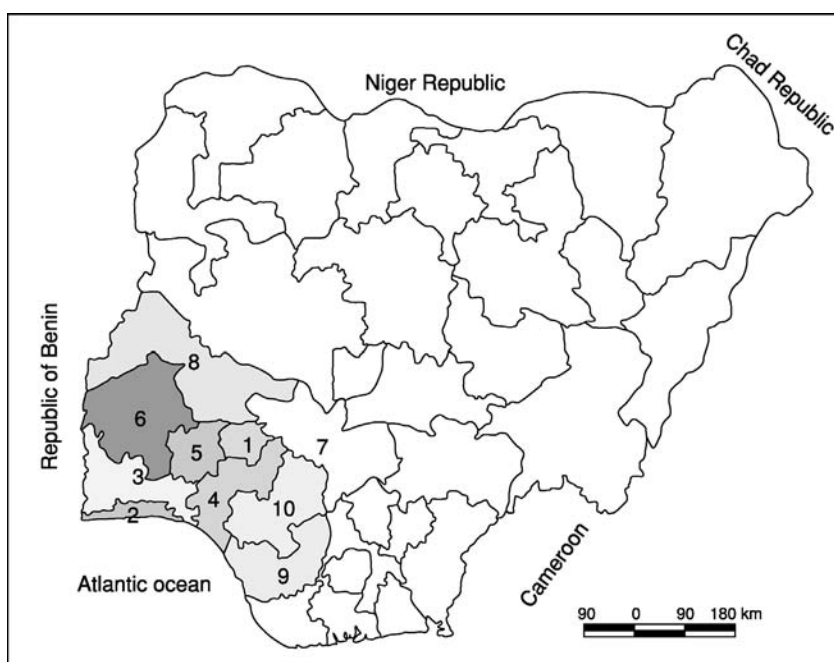


Figure 1 Nigerian states in which Yoruba is spoken as a first or second language. Key to states: 1, Èkìtì; 2, Lagos; 3, Ògùn; 4, Òndó; 5, Òşun; 6, Òyó; 7, Kogi; 8, Kwara; 9, Delta; 10, Edo.

Syntactic Characteristics

Noun Classes, Gender System, and Number Yoruba has no noun class or grammatical gender. There are no separate noun forms to distinguish singular from plural. However, when necessary, plurality can be marked by using some (pro)nominal forms before nouns, or by using certain demonstratives, as well as by reduplicating adjectives after nouns.

Possessive Noun–Noun Constructions In a sentence, the second noun in a possessive noun–noun construction can be focused by front shifting it, with its original position being occupied by the appropriate pronoun qualifier. After certain verbs, it is also possible to permute the constituent nouns with the particle *ní* obligatorily intervening (see Owolabi, 1976: 40–43).

Past and Present Actions Action verbs generally convey past action, whereas stative verbs convey past or present action.

Verbal Constructions There are verbal constructions in which subject and object nouns can switch positions with little or no difference in meaning, and verbal constructions in which the verbs are repeated after their objects; in addition, some verbal constructions contain verbs that are used for asking questions, verbs that are negative in meaning, or verbs that obligatorily select the particle *ní* (see Awobuluyi, 1978: 53–62).

Morphological Characteristics

In order to form morphologically complex nouns, certain prefixes (for example, *à-*, *è-*, *o-*, *i-*, and *ài-*) that are usually attached to roots that are verbs or verb phrases, and sometimes to ideophonic adjectives, cannot be combined. However, the prefixes *oní-* and *oni-*, which are attached to nouns or noun phrases, can be combined and can also combine with the former class of prefixes, resulting in nouns that denote emphasis (see Owolabi, 1995: 93–102).

Phonetic/Phonological Characteristics

Vowel Co-occurrence Restrictions and Vowel Elision Yoruba operates a partial system of vowel harmony in which the set /e, o/ mutually excludes the set /ɛ, ɔ/ in polysyllabic underived words. Also, in words with a vowel₁-consonant-vowel₂ (V₁-C-V₂) pattern, neither the nasalized vowels nor /u/ can occur as V₁. Vowel elision, resulting in contractions, is quite erratic apart from the relatively predictable elision of the vowel /i/,

| Consonants | | | | | | | |
|------------|---|---|---|-----------|---|----|----|
| b | t | d | ɖ | k | g | kp | gb |
| f | s | ʃ | | | | | h |
| m | n | r | | | | | |
| | l | | j | | w | | |
| Vowels | | | | | | | |
| Oral | | | | Nasalized | | | |
| i | | u | | ĩ | | ũ | |
| e | | o | | ẽ | | õ | |
| ɛ | | ɔ | | | ã | | |

Figure 2 Phonetic groupings of Yoruba consonants and vowels. Tones are indicated by diacritical marks: high (ˊ), mid (ˋ), and low (ˌ).

the initial vowel of the second noun of noun–noun combinations, or the vowels of the standard forms of words in combination with the dialectal forms (see Bamgboṣe, 1989). **Figure 2** shows the phonetic groupings of Yoruba consonants and vowels; tones are indicated by diacritical marks: high (ˊ), mid (ˋ), and low (ˌ).

The Assimilated Low Tone In addition to the high, mid, and low tones in Yoruba, a tonal feature referred to as ‘the assimilated low tone’ occurs when a low tone disappears in certain contracted expressions or in some single polysyllabic words, but its influence is still felt on the following syllable (see Bamgboṣe, 1966).

Restriction on the Occurrence of the High Tone The high tone never occurs on V₁ in words of V₁-C-V₂ pattern.

Tones of Pronouns With the exception of the second-person plural pronoun object, the lexical tones of the verbs determine the tones of all pronoun objects. Similarly, the tones of some subject pronouns vary before the verbal particle *ní*.

Other Characteristics

Various semantic effects (e.g., emphasis, anger, and anxiety) can be achieved by employing the devices of reduplication, prefixation, and vowel lengthening, or by using ‘intensifiers.’ Focusing whole sentences

apart from sentence constituents is also common. The following sentence provides an example of the Yoruba language:

Alákàá ni ilé è wó, tí olè sì kó mí ní èrù, sùgbón ó dún mí pé Olú jó, jó, jó ni léyìn iṣẹ̀lẹ̀ wònyí, èyí tó mú kí àwọn ọ̀lómòwé ọ̀rẹ̀ Olú rò pé kò fẹ̀ èmi àti Alákàá fẹ̀ ọ̀rọ̀, àmò láipé, Alákàá yòò rálẹ̀ tuntun, yòò sì ròkò pèlú.

It was Alákàá's house that fell while thieves stole my property, but it pained me that what Olú did was to really dance after these incidents, which made Olú's educated friends think that he isn't happy to see Alákàá and me prosper, however, Alákàá will soon purchase a new house and a vehicle as well.

Note that *è* occupies the original position of the front-shifted noun *Alákàá*, and that *mí ní èrù* is a possessive construction resulting from permutation. In *Alákàá* (from *'oní àkà'*), the influence of the assimilated low tone is felt; *mí* has a high tone after the low-tone verb *dún*, but a mid tone after the high-tone verb *kó*. For emphasis, *jó* is reduplicated and *Alákàá* and the phrase beginning with *sùgbón* are focused by placing the focus marker *ní* after them. Plurality is indicated by *wònyí* and *àwọn*, *ọ̀lómòwé* comprises the verb phrase *mòwé* and the prefixes *ò-* and *oní-*, and *fẹ̀* is repeated after *èmi àti Alákàá*. The 'a' of *rà* is retained in the verb-noun contraction *rálẹ̀* (from *ra ilé* 'purchase a house') but is elided in *ròkò* (from *ra ọ̀kò* 'purchase a vehicle'). A phonemic transcription of the sentence is as follows:

álá'ká lí ilé è wó, tí olè sì kó mí lí èrù, jùgbón ó dǔn mí kpé olú jó jó jó lí lẹ̀jì iṣẹ̀lẹ̀ wòyí, èyí tó mú kí àwǒ ọ̀lòmòwé ọ̀rẹ̀ olú rò kpé kò fẹ̀ èmi àti álá'ká fẹ̀ ọ̀rò, àmǒ láikpé, álá'ká jóò rálẹ̀ tǔtǔ, jóò sì ròkò pèlú.

Other Points of Relevance

The Yoruba language comprises about 20 dialects. There is also a form of Yoruba popularly referred to as Standard Yoruba. In all of the Nigerian states in which Yoruba is spoken natively, the Standard Yoruba and the Yoruba dialects are spoken. However, a diglossic situation exists where the Standard Yoruba is the high variety and the dialects are the low variety, although the use of some of the dialects in publications, broadcasts, and native courts (in particular) is increasing. The variety of Yoruba described in this article is Standard Yoruba. Yoruba vocabularies occur in poetic recitations associated with rituals and cults in Brazil as well as in Sierra Leone, where the influence of Yoruba is also felt in Krio loanwords and personal names.

Yoruba is one of the three major languages in Nigeria (the other two are Hausa and Igbo). It is

taught as a subject at the primary, secondary, and tertiary levels. At least eight universities in Nigeria offer first degree and/or higher degree courses in Yoruba. Literature in the language is also very extensive. According to government policy, in the states in which Hausa, Igbo, or Yoruba is not a mother tongue, one of these three languages is a compulsory subject at secondary school level. The three languages are also to be used in the National Assembly in addition to English, and one of the state assemblies in the Yoruba-speaking states is currently using Yoruba in the same way. Similarly, the Nigerian federal government has embarked on the translation of the 1999 constitution of the Federal Republic of Nigeria into Yoruba and the other two major Nigerian languages (Hausa and Igbo), in order to facilitate the usage of these languages in the domain of legislation.

A Yoruba metalanguage (available in two volumes) has facilitated the use of Yoruba for writing textbooks and as a medium for teaching the language at all levels of education, whereas the Six-year Primary Project at the Obafemi Awolowo University (formerly the University of Ife), in Ile-Ife, Oṣun State, aims at demonstrating that all subjects can be taught in Yoruba at the primary level.

In the neighboring Republic of Benin, the Yoruba, Aja, and Fon languages are studied at the university level; Yoruba was adopted (along with Aja, Fon, Bariba, Bendi, and Waama) as an official language of the National Assembly in 1983 (see Adeniran, 2004: 437, 442).

Bibliography

- Adeniran W (2004). 'Language attitudes in a francophone city: the case of the Yoruba of Porto-Novo.' In Owolabi K & Dasylyva A (eds.). 437–460.
- Akinkugbe O O (1978). 'A comparative phonology of Yoruba dialects, Isekiri, and Igala.' Ph.D. thesis, Ibadan: University of Ibadan.
- Awobuluyi O (1978). *Essentials of Yoruba grammar*. Oxford: Oxford University Press.
- Bamgboṣe A (1966). 'The assimilated low tone in Yoruba.' *Lingua* 16(1), 1–13.
- Bamgboṣe A (1989). When rules fail: the pragmatic of vowel elision in Yoruba. Paper presented at the Linguistic Association of Nigeria conference, University of Jos.
- Bendor-Samuel J (ed.) (1989). *The Niger-Congo languages*. Lanham, MD: University Press of America.
- Capo H B C (1989). 'Defoid.' In Bendor-Samuel J (ed.). 275–290.
- Hair P E H (1967). *The early study of Nigerian languages. West African language monographs 7*. Cambridge: Cambridge University Press.

- Owolabi D K O (1976). 'Noun-noun constructions in Yoruba: a syntactic and semantic analysis.' Ph.D. thesis, Ibadan: University of Ibadan.
- Owolabi D K O (1993). 'Yoruba.' In *Encyclopedia of language and linguistics*, 1st edn. Oxford: Pergamon Press Ltd. 5074–5076.
- Owolabi K (1995a). 'More on Yoruba prefixing morphology.' In Owolabi K (ed.), 92–112.
- Owolabi K (ed.) (1995b). *Language in Nigeria: essays in honour of Ayo Bamgboṣe*. Ibadan: Group Publisher.

- Owolabi K (2004). 'On the translation of the 1999 Constitution of the Federal Republic of Nigeria into selected Nigerian languages.' In Owolabi K & Dasylyva A (eds.), 523–537.
- Owolabi K & Dasylyva A (eds.) (2004). *Forms and functions of English and indigenous languages in Nigeria: a festschrift in honour of Ayo Banjo*. Ibadan: Group Publishers.
- Williamson K (1989). 'Benue-Congo overview.' In Bendor-Samuel J (ed.), 3–46.

Yukaghir

G D S Anderson, Max Planck Institute, Leipzig, Germany and University of Oregon, Eugene, OR, USA

© 2006 Elsevier Ltd. All rights reserved.

Yukaghir is not a single language, but is actually a small language family consisting of two nearly extinct languages of northeastern Siberia, i.e., Tundra (Northern Yukaghir) and Kolyma (Southern Yukaghir). The speakers of these languages and the languages themselves are known by the self-designations of *Wadul* (Tundra) and *Odul* (Kolyma). Fewer than 200 total speakers, possibly as few as 40, live in northeastern Siberia. Conventionally labeled as language isolates, some consider Tundra and Kolyma to be distant relatives of the Uralic languages. The Yukaghiric family probably originally included two now extinct languages, Chuvan and Omok.

Yukaghiric-speaking peoples were once dominant over a vast area in northeastern Siberia, practicing reindeer husbandry and subsistence hunting and fishing. Yukaghiric-speaking peoples at first assimilated Tungusic-speaking peoples culturally, but eventually the Even (Tungusic) people assimilated the Yukaghiric people linguistically, and now include a discernible Yukaghiric substrate. The Tundra Yukaghir language is spoken in two villages, Andryushkino and Kolymaskoe. Kolyma Yukaghir is found predominantly in the village of Nelemnoe. Both Yukaghir languages are moribund, spoken now only by a few elders. The Yukaghir people have shifted mainly to speaking Russian, but in Andryushkino village they are mostly shifting to Yakut (Sakha), a locally dominant Turkic language.

Yukaghir possesses a range of contrastively palatalized segments in the consonant system, a pattern commonly found throughout the Siberian area. Unlike most northern and Siberian languages, Yukaghir is like Yakut in not permitting *ŋ*-in word-initial

position. However, the common four-way place contrast for nasals (*m/n/ɲ/ŋ*) seen across the languages of Siberia is an old and stable feature in Yukaghir, going back at least to the Proto-Yukaghir(ic) stage (Anderson 2003). Example (1) shows word forms in Tundra, Kolyma, and Proto-Yukaghir (from Krejnovich, 1958, 1982: 13–14):

| (1) Tundra | Kolyma | Proto-Yukaghir | Gloss |
|------------|--------|----------------|---------------|
| nonol | nonol | *nonol | 'loop, noose' |
| amun | amun | *amun | 'bone' |
| ñaka | ñaka | *ñaka | 'together' |
| aŋa-ŋ | aŋa | *aŋa | 'mouth' |

Like most other Siberian languages, Yukaghir makes use of a range of case forms of nouns. This includes both areally common and typologically unusual formations. To the areally common group of features belong the opposition of instrumental (INSTR) case forms (Example (2a); Krejnovich, 1982: 49–50)) and comitative (COM) case forms (Examples (2b) and (2c); Krejnovich 1982: 45, 46)):

| (2a) Tundra Yukaghir | Kolyma Yukaghir |
|----------------------|-----------------|
| -lek, -leŋ | -le, -lek |
| sā-lek pajduk | čoyoye-le |
| 'hit with a stick' | 'with a knife' |

In Examples (2b) and (2c), comitative denotes possession as well, and conjoins two nouns (PV, preverb):

| (2b) Kolyma Yukaghir | |
|---|---------------------|
| núme-ñej | |
| dwelling-COMIT | |
| 'with a dwelling,' 'he has a dwelling' | |
| (2c) Tundra Yukaghir | |
| ile-ñej | ila:me me-qaldej-ni |
| reindeer-COMIT | dog PV-run.off-3PL |
| 'the dog ran off with the reindeer,' 'the dog and the reindeer ran off' | |

To the unusual group of case features belongs the characteristically Yukaghir but typologically unusual system of 'focus' marking. Simplifying matters

somewhat, this system is as follows: there is (1) an unmarked form that encodes speech act participants and lexical nouns in agent-focus and agent/subject-topic functions; (2) a marked 'neutral' case used with nonlocutor agent/subject topics, object-topics with locutor agents, and forms that lack the focus case (nonlocutor personal pronouns, proper names, etc.); and (3) the 'focus' case that encodes subject or object focus. Noun phrases (NPs) marked with focus frequently reference indefinite NPs, and focus-case marking often serves to introduce participants into the discourse (see Maslova (2003a: 51ff) for further details).

In terms of the clausal syntax of simple and complex sentences, Yukaghir is similar to many other indigenous Siberian languages. The language shows dominant subject-object-verb constituent order and uses a wide range of adverbial 'converb' forms in subordinate clause formation, as well as the characteristic system of case-marked nominalized verbs to mark a wide range of functional subtypes of subordinate clauses, as shown in Example (3a) (the first two words are from Krejnovich (1958: 198) and Fortescue (1988: 41), respectively; the original source for the third word is unknown) and Example (3b) (from Maslova (2003b: 372)) (abbreviations: NOM, nominal; LOC, locative; POSS, possessive; PL, plural; NF, nonfinite; NEG, negation; PROHIB, prohibitive; ACC, accusative; DESID, desiderative; INTRANS, intransitive):

(3a) Yukaghir

| | | |
|-----------------|---------------|---------------|
| u:r-eŋ | u:-l-rane | u:-l-lek |
| go:ACTION. | go-ACTION. | go-ACTION. |
| NOM-LOC | NOM-LOC.II | NOM-INS |
| 'when (I) went' | when he went' | 'after going' |

(3b) Kolyma Yukaghir

| | |
|---|-----------------|
| qa:qa:-pe-gi | ajli-de-ge |
| grandfather-PL-POSS | forbid-3.NF-LOC |
| "el+qon-ti-lek" | mon-de-ge |
| NEG-go-PL-PROHIB | say-3.NF-LOC |
| tamun-gele | uørpe-p-ki |
| that-ACC | child-PL-POSS |
| el+med-o:l-ti | |
| NEG-listen-DESID- | |
| 3pl.intrans | |
| 'Their grandfather forbids (it), saying "don't go" but the children do not obey'. | |

Bibliography

- Anderson G D S (2001). 'Deaffrication in the (Central) Siberian Area.' In Aronson H I (ed.) *Non-Slavic languages 9: linguistic studies*. Columbus, OH: Slavica. 1-17.
- Anderson G D S (2003). 'Towards a phonological typology of Native Siberia.' In Holisky D A & Tuite K (eds.) *Current trends in Caucasian, East European and*

- Inner Asian linguistics. Papers in honor of Howard I. Aronson*. Amsterdam & Philadelphia: John Benjamins. 1-22.
- Angere J (1956). *Die uralo-jukagirische Frage. Ein Beitrag zum Problem der sprachlichen Urverwandtschaft*. Uppsala-Stockholm: Almqvist and Wiksell.
- Angere J (1957). *Jukagirisch-Deutsches Wörterbuch*. Wiesbaden: Almqvist and Wiksell.
- Bouda K (1941). 'Die finnisch-ugrisch-samojedische Schicht des Jukagirischen.' *Ungarische Jahrbücher* 17, 80-101.
- Collinder B (1940). 'Jukagirsch und Uralisch.' *Uppsala Universitets Årsskrift* 8, 1-143.
- Comrie B (1992). 'Focus in Yukagir (Tundra dialect).' In Aronson H I (ed.) *The non-Slavic languages of the USSR*. Chicago: Chicago Linguistic Society. 55-70.
- Fortescue M (1988). 'The Eskimo-Aleut-Yukagir relationship: an alternative to the genetic/contact dichotomy.' *Acta Linguistica Hafniensia* 21(1), 21-50.
- Fortescue M (1996). 'Grammaticalized focus in Yukaghir. is it really grammaticalized and is it really focus?' In Engberg-Pedersen E *et al.* (eds.) *Content, expression and structure. Studies in Danish functional grammar*. Amsterdam: Benjamins. 17-38.
- Jochelson W (1900). *Materialy po izučeniju jukagirskogo jazyka i fol'klora*. Sankt-Peterburg: Akademija Nauk.
- Joxel'son V I (1934). 'Odul'skij (jukagirskij) jazyk.' In Krejnovich E A (ed.) *Jazyki i pis'mennost' Sibiri III*. Moscow: Gosudarstvennoe učebno-pedagogičeskoe izdatel'stvo. 149-180.
- Krejnovich E A (1955). 'Sistema morfoložičeskogo vyrazhenija logičeskogo udarenija v jukagirskom jazyke.' *Institut Jazykoznanija. Doklady i Soobščhenija* 7, 99-115.
- Krejnovich E A (1958). *Jukagirskij jazyk*. Moscow/Leningrad: Akademija Nauk SSSR.
- Krejnovich E A (1982). *Issledovanija i materialy po jukagirskomu jazyku*. Leningrad: Nauka.
- Kurilov G N (1991). *Jukagir-Russian dictionary*. Yakutsk: Jakutskoe Knizhnoe izdatel'stvo.
- Maslova E (1989). 'Retsiprok v jukagirskom jazyke.' *Sowjetische Finnisch-Ugrische Sprachwissenschaft* XXV(2), 120-127.
- Maslova E (1993). 'The causative in Yukagir.' In Comrie B & Polinsky M (ed.) *Causatives and transitivity*. Amsterdam: Benjamins. 271-285.
- Maslova E (1997). 'Yukagir focus system in a typological perspective.' *Journal of Pragmatics* 27, 457-475.
- Maslova E (ed.) (2001). *Jukaghir texts. Tunguso-Sibirica series* (vol. 7). Wiesbaden: Harrassowitz.
- Maslova E (2003a). *Tundra Yukaghir. Languages of the world/materials* 372. Munich: Lincom.
- Maslova E (2003b). *A grammar of Kolyma Yukaghir*. Berlin: Mouton de Gruyter.
- Nikolaeva I A (1988). 'O sootvetsvijax uralskix affrikat i sibilantov v jukagirskom jazyke.' *Sovetskoe Finno-Ugrovedenie* 2, 81-85.
- Nikolaeva I (ed.) (1989). *Fol'klor Jukagirov verxnej Kolymy* (vols. 1-2). Yakutsk: Yakutskij Gosudartvennyj Universitet.

- Nikolaeva I (ed.) (1997). *Yukaghir texts. Specimina Sibirica* 13. Szombathely: Savariae.
- Nikolaeva I (2000). *Chresthomatia Yukagirica*. Budapest: ELTE BTK, Finnugor Tanszék.
- Nikolaeva I A & Xelimskij E (1997). 'Jukagirskij jazyk.' In *Jazyki narodov mira. Paleoaziatskie jazyki*. Moscow: Indrik. 155–168.
- Ostrowski M (1983). *Zur Nomen:Verb Relationierung im Wogulischen, Jurakischen und Yukagirischen*. Köln: Arbeiten des Kölner Universalien-Projekts.
- Spiridonov V (1997). *Russian-Yukagir dictionary*. Zyrianka: Jakutskoe Knizhnoe izdatel'stvo.
- Tailleur O G (1959). 'Les unique donnees sur l'omok, langue éteinte de la famille youkaghire.' *Orbis* VIII, 78–108.
- Tailleur O G (1962). 'Le dialecte tchouvane du Youkaghir.' *Ural-Altaische Jahrbücher* XXIV, 55–99.
- Tailleur O G (1965). 'La flexion personnelle en Youkaghir.' *Études Finno-Ougriennes* 2, 67–88.
- Vakhtin N (1992). *The Yukagir language in sociolinguistic perspective. Monograph series* 2. 1991. Steszew: International Institute of Ethnolinguistic and Oriental Studies.
- Veenker W (1989). *Tundrajukagirisches Wörterverzeichnis*. Hamburg: Helmut Buske.

Z

Zapotecan

G A Broadwell, State University of New York, Albany, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Zapotecan languages belong to the Otomanguean family and are spoken across a large portion of Oaxaca, Mexico. Zapotecan is divided into two branches – Chatino and Zapotec proper. The number of languages in each group is controversial, but the Summer Institute of Linguistics recognizes 6 Chatino languages and 58 Zapotec languages.

Zapotec languages are spoken by over 200 000 people located over much of the eastern half of Oaxaca. Varieties of Zapotec divide broadly into three groups. The Valley-Isthmus group includes most varieties spoken in the valley of Oaxaca, extending to the Isthmus of Tehuantepec. The Northern group is spoken in the mountains to the north of the Valley of Oaxaca, and the Southern group is spoken in the mountains to the south. All three groups are quite diverse and contain many distinct languages.

Chatino languages are spoken in a smaller mountainous area in southwestern Oaxaca by perhaps 30 000 people.

The earliest documentation of Zapotecan languages comes from grammars, dictionaries, and religious material from the Spanish colonial period. Archaeological sites associated with the Zapotec state of ca. 100 B.C. to ca. 900 A.D. also contain Zapotec hieroglyphic writing. Efforts to decipher Zapotec hieroglyphics are still ongoing.

Phonological Characteristics

In Zapotec languages, most consonants are divisible into two morphophonologically defined groups called ‘fortis’ and ‘lenis.’ For stops and fricatives, fortis is largely equivalent to voiceless and lenis to voiced. However, the fortis/lenis distinction is also found in the nasals and sonorants, where the phonetic realization of fortis is not voicelessness, but some other characteristic that generally includes longer duration.

The morphological examples in (2) below from Mitla Zapotec show that long sonorants and voiceless obstruents seem to form a natural class. Many analyses of Zapotec historical phonology also analyze fortis consonants as having originated from geminates and consonant clusters.

Many Zapotec languages, especially those in Valley group, show contrastive phonation type differences. San Dionicio Ocotepec Zapotec, for example, shows a distinction between modal, breathy, creaky, and checked vowels. Consider the following minimal and near-minimal pairs:

- (1) San Dionicio Ocotepec Zapotec
- | | | |
|----------|------------|-----------|
| [bɛ̞] | ‘flame’ | (breathy) |
| [bɛ̞] | ‘meat’ | (creaky) |
| [bá:ld] | ‘how many’ | (modal) |
| [bâ] | ‘bullet’ | (breathy) |
| [bɛ̞ld] | ‘fish’ | (breathy) |
| [bɛ̞ʔld] | ‘snake’ | (checked) |

All Zapotecan languages are tonal. The number of reported tones varies from two tonal levels to four, with a variety of contour tones. The largest number of tonal contrasts in the family appears to be Zoogocho Zapotec, which is reported to have four tone levels and seven contours, for a total of 11 tonal contrasts.

Morphological Characteristics

Zapotec languages do not have a passive, but generally show morphological relationships between pairs of verbs that differ in valence. In one typical pattern, an intransitive stative verb begins with a lenis consonant, while the corresponding transitive active verb begins with the corresponding fortis consonant. Consider the following examples from Mitla Zapotec:

- (2) Mitla Zapotec
- | | | | |
|-------|-------------------|--------|-----------------|
| [zæb] | ‘to sink (intr.)’ | [sæb] | ‘to sink (tr.)’ |
| [dɛb] | ‘to be wrapped’ | [tɛb] | ‘to wrap’ |
| [nɪt] | ‘to be lost’ | [nnɪt] | ‘to lose’ |
| [lib] | ‘to be tied’ | [llib] | ‘to tie’ |

There are generally also other such pairs that show less regular correspondences.

Zapotec verbs are inflected with aspectual prefixes. The number of aspects varies from language to language. In Zoogocho Zapotec, for example, there are continuative, stative, completive, potential, and dubitative aspects. San Lucas Quiaviní Zapotec has progressive, habitual, perfective, irrealis, subjunctive, neutral, and definite (future) aspects.

Zapotec verbs do not show agreement, though pronominal subjects (and some pronominal objects) cliticize to the verb. Consider the following examples from San Dionicio Ocotepéc Zapotec:

(3) San Dionicio Ocotepéc Zapotec

(3a) Û-dàw rée = bíny bzyáá.
COM-eat PLUR = person bean
'The people ate beans.'

(3b) Û-dàw = réby bzyáá.
COM-eat = 3.HUMAN.PLUR bean
'They ate beans.'

(3c) Û-dàw = réby = rênny
COM-eat = 3.HUMAN.PLUR = 3.INAM.PLUR
'They ate them.'

Zapotec languages often show a large number of third-person pronominal categories. For example, San Lucas Quiaviní Zapotec distinguishes between proximal (near) and distal (far) third persons, as well as between animal, respectful, formal, and reverential third-person categories. These categories are not morphologically marked on nouns themselves, but on independent or clitic pronouns that are coreferential with the nouns. Pronominal category is also not completely fixed, but may vary somewhat according to a speaker's point of view and according to the structure of the narrative.

Syntactic Characteristics

Zapotecan languages show head-initial order in phrases. Clauses are VSO, noun phrases are N-initial, and the language is prepositional. The following examples from San Dionicio Ocotepéc Zapotec and Yaitepec Chatino show these properties:

(4) San Dionicio Ocotepéc Zapotec
Û-dèèdy Gùstááb j-kèès Màrîi lòò Móòny
COM-give Gustavo POSS-pot Maria to Ramón
'Gustavo gave Maria's pot to Ramón.'

(5) Yaitepec Chatino (Pride, 1965: 82)
Njî?yú³² ne?³ yka³ lo?o¹ ta?a²³ lo?o¹
cutting he wood with brother with
siyera⁴ ka³ sî² bra³kô?².
saw yesterday evening then
'He was cutting wood with his brother with a saw yesterday evening then.'

Though the languages are verb-initial, there is generally an elaborated hierarchy of positions to the left

of the verb. These frequently include special positions for topical, focal, negative, and interrogative phrases. The following examples from Quiegolani Zapotec illustrate several of these positions:

(6) Quiegolani Zapotec

[Tʂu]_{Interrog} [men]_{Neg} wii-t?
who nothing saw-NEG
'Who saw nothing?'

(7) Quiegolani Zapotec

[Laad ʂ-unaa dolf]_{Focus} ɖe
FOC POSS-wife Rodolfo already
z-u nga.
PROG-stand there
'Rodolfo's wife was already standing there.'

All the Zapotecan languages also appear to show the phenomenon known as 'pied-piping with inversion.' When a subpart of NP or PP (and sometimes other constituents) is questioned, the entire phrase moves to the clause-initial interrogative position, but shows an inverted order, in which the interrogative precedes the head of the phrase:

(8) San Dionicio Ocotepéc Zapotec

[Túú lòò]_{Interrog} ù-dèèdy Gùstááb gèès?
who to COM-give Gustavo pot
'Who did Gustavo give the pot to?'

In broad syntactic terms, Zapotecan languages generally conform to the areal features of other Mesoamerican language families such as Mayan, Mixe-Zoquean, and Totonacan. Zapotecan languages differ from these other groups in lacking agreement and voice morphology and having a more rigid word order.

Conclusion

All Zapotecan languages are endangered, and in some communities, only a few elders speak the language. In other communities, the language is spoken by a much larger proportion of the population, but there are still economic pressures that favor language shift to Spanish or emigration to the United States and other parts of Mexico. These factors make language preservation and documentation work an urgent priority.

Bibliography

- Black C (2000). *Quiegolani Zapotec syntax*. Dallas, TX: Summer Institute of Linguistics.
Briggs E (1961). *Mitla Zapotec grammar*. Mexico City: Instituto Lingüístico de Verano.
Broadwell G A (2001). 'Optimal order and pied-piping in San Dionicio Zapotec.' In Sells P (ed.) *Formal and empirical issues in optimality theoretic syntax*. Stanford: Center for the Study of Language and Information Publications. 197–123.

- Long R & Cruz S (1999). *Diccionario zapoteco de San Bartolomé Zoogocho, Oaxaca*. Coyoacán, Mexico: Instituto Lingüístico de Verano.
- Munro P & Lopez F (1999). *Di'csyoonarry X:tèè'n Dii'zh Sah Sann Lu'uc: San Lucas Quiavini Zapotec Dictionary*.

- Los Angeles: UCLA Chicano Studies Research Center Publications.
- Pride K (1965). *Chatino syntax*. Norman, OK: Summer Institute of Linguistics and University of Oklahoma.

Zulu

A van der Spuy, University of the Witwatersrand, Johannesburg, South Africa

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Zulu, also known as isiZulu, is a Southern Bantu language, and is one of the 11 official languages of South Africa. With over nine million speakers, it is one of the country's major languages, and is used in broadcasting, journalism, and the national and provincial parliaments. Famous as the language of the Zulu empire of the 19th century, it has a growing literature, and there are efforts to develop a technical vocabulary for use in the teaching of mathematics and other sciences.

The language has been the subject of a considerable number of grammatical and linguistic studies, dating back to works of 19th-century pioneers such as Grout (1859). Zulu is closely allied to Xhosa, Swati, and Ndebele, and there is a high degree of mutual intelligibility between these languages, to the extent that it could be argued that they are all varieties of one language, Nguni. The findings of linguistic studies of the other Nguni languages are very frequently applicable to Zulu as well.

Morphology

Zulu displays the typical Bantu morphological features: it is highly agglutinative, and its nouns are divided into various classes, which command distinctive agreement morphology (see 'Syntax' below). Most of the noun classes occur in singular/plural pairs, for example, a noun such as *inja* 'dog' (class 9) will have a plural in class 10, *izinja* 'dogs.' Older studies classified the noun classes according to this pairing (e.g., Doke, 1927), but Canonici (1990) has proposed a classification according to agreement characteristics. From this point of view there are 12 noun classes. There is an elaborate tense and aspect system, and verbs may take valency-changing suffixes (known as 'extensions,' e.g., causative *-is-* in *fund-is-a* 'cause to learn, teach'; passive *-w-* in

fund-w-a 'be learnt'; reciprocal *-an-* in *fund-is-an-a*, 'teach each other').

The morphology of the language and the semantics of the various grammatical forms have been the focus of linguistic research into Zulu for the past 80 years. Dominating most studies has been Doke's model (1927), which sought to describe Bantu languages in terms appropriate to that family, rather than according to established Latin, Greek, or English terminology. Subsequent accounts have largely been refinements of the Dokean model, e.g., Cope (1984) and Poulos and Msimang (1998).

Phonology

Zulu phonology has been described in a number of works, most notably and comprehensively in Khumalo (1987). Like many Bantu languages, it has an (N)CV syllable structure. There are 40 phonologically distinct consonants, and five vowels. Vowel length is usually predictable, but occasionally distinctive (for example, between the remote past tense and the past consecutive tense: *wa:hamba* 'he/she walked,' *wahamba* 'and then he/she walked'). There is a stricture on the occurrence of two vowels in juxtaposition at surface level, which leads to rules of vowel merging; for example, possessive *a-* prefixed to *inja* 'dog' yields *enja* 'of the dog.' Other processes that have been frequently studied include palatalization and so-called nasal strengthening, where nasals in N + C clusters change the nature of the following C. For example, an aspirated consonant in this position will lose its aspiration, so that the root *phil-* 'live' becomes *-pil-* in the class 9 noun *impilo* 'life.' The language has a high/low tone contrast, and a (derived) high-low tonal cluster may occur on bimoric vowels.

Syntax

Zulu has a basic SVO word order. Relative clauses and possessive phrases follow the head noun, and auxiliaries precede the verb. There is considerable agreement marking, as in the following example,

where the affixes glossed as AGR all agree with the class 9 noun *moto* 'car.' ('REL' stands for 'relative marker'.)

| | | |
|---------------------------------------|------|-----------------|
| Le-y-o | moto | e-n-tsha |
| DEM-AGR-that | car | REL-AGR-new |
| o-yi-theng-ile | | i-fik-ile. |
| REL.PERS2-AGR-buy-PERF | | AGR-arrive-PERF |
| 'that new car you bought has arrived' | | |

Formal linguistic studies of syntactic phenomena in the Southern Bantu languages have frequently been cast in terms of the Chomskyan Principles and Parameters framework. Much of this work has concentrated on Xhosa rather than Zulu, e.g., Du Plessis and Visser (1992). Little or no work has been done in other frameworks such as Head-Driven Phrase Structure Grammar or Lexical Functional Grammar, although the latter has proved useful in descriptions of other Bantu languages such as Chewa (Nyanja).

Historical and Comparative Linguistics

Like other languages in the east of the Bantu area, Zulu shows fricativization of stops before the 'extra high' vowels of proto-Bantu, which subsequently merged with the high vowels. There is no synchronic operation of Meinhof's law or Dahl's law, and the verbal suffixes (extensions) show no vowel harmony. The noun classes found in the language are (to use the numbering system by which they are known in Bantu studies) 1–7, 9–11, 14, and 15 (the last being used for the infinitive). Only fossilized versions of locative classes remain as adverbs, such as *phansi* 'below.' Unusually for a Bantu language, Zulu has noun suffixes, e.g., the feminine marker *-kazi*. Several of the original Bantu verbal suffixes survive only in unproductive forms (e.g., *-ul-* in words such as *khumbula* 'remember'). The language is well known for its extensive borrowing of Khoi and San words and sounds, noticeably the click consonants in words such as *iqhwa* 'snow.' It has also incorporated many words from Afrikaans and English.

Sociolinguistics

Zulu has certain marked speech forms which are of interest to sociolinguists. An example is *hlonipha*, the speech form traditionally used by married women,

who have to avoid words that sound like the names of any of their close male in-laws, and therefore acquire a radically altered vocabulary (see Herbert, 1990). Another example is *isicamtho*, a group-marking variety used by young urban men, which has borrowed many words from English, often with radical change of meaning. There are several distinctive dialects of Zulu, and in some urban varieties the boundaries between Nguni languages have become less marked. In urban areas there is also much code switching between Zulu and the Sotho languages, and between Zulu and Afrikaans or English.

Current Research Directions

Zulu has long been one of the most studied Bantu languages, and it remains a focus of much research in the areas discussed above, and also in new directions including child language acquisition (Suzman, 1996) and computational linguistics (Bosch and Pretorius, 2002).

Bibliography

- Bosch S E & Pretorius L (2002). 'The significance of computational morphological analysis for Zulu lexicography.' *South African Journal of African Languages* 22, 11–20.
- Canonici N (1990). 'Noun classes and subclasses.' *South African Journal of African Languages* 10, 52–58.
- Cope A T (1984). 'An outline of Zulu grammar.' *African Studies* 43, 83–104.
- Doke C M (1927). *Textbook of Zulu grammar*. Johannesburg: Longman.
- Du Plessis J A & Visser M W (1992). *Xhosa syntax*. Pretoria: Via Afrika.
- Grout L (1859). *The isiZulu: a grammar of the Zulu language*. Pietermaritzburg: May and Davis.
- Herbert R K (1990). 'Hlonipha and the ambiguous woman.' *Anthropos* 85, 455–473.
- Khumalo J S M (1987). An autosegmental account of Zulu phonology. Ph.D. thesis, University of the Witwatersrand.
- Poulos G & Msimang C T (1998). *A linguistic analysis of Zulu*. Cape Town: Via Afrika.
- Suzman S M (1996). 'Acquisition of noun class systems in related Bantu languages.' In Johnson C & Gilbert J H (eds.) *Children's language* 9. Hillsdale, NJ: Lawrence Erlbaum Associates. 87–104.

INDEX

NOTES

Cross-reference terms in italics are general cross-references, or refer to subentry terms within the main entry (the main entry is not repeated to save space). Readers are also advised to refer to the end of each article for additional cross-references - not all of these cross-references have been included in the index cross-references.

The index is arranged in set-out style with a maximum of three levels of heading. Major discussion of a subject is indicated by bold page numbers. Page numbers suffixed by T and F refer to Tables and Figures respectively. *vs.* indicates a comparison.

Subentries (or subsubentries) to a specific index entry having the same page number, have been included to indicate the breadth of the discussion (as opposed to just the location), as additional assistance to the reader.

This index is in letter-by-letter order, whereby hyphens and spaces within index headings are ignored in the alphabetization. Prefixes and terms in parentheses are excluded from the initial alphabetization.

A

- Aasen, Ivar Andreas, Norwegian 785
 Abbey 631
 see also Kwa languages
 Abé 631
 see also Kwa languages
 Abhidhamma Pitaka 831–832
 Abidji 631
 see also Kwa languages
 Abkhaz 1–2
 case markers 1–2
 classification 251
 grammar 1
 history 1
 influence from other languages 2
 orthography 1
 phonemes 1
 potential/involuntary constructions 2
 preverbal grade systems 2
 relative strategy 2
 Stative-Dynamic opposition 2
 use of 1
 verbal complexity 2
 see also Caucasian languages;
 Georgian; North Caucasian
 languages
 Abkhazia
 Armenian 68
 Caucasian languages 192
 Abnaki
 classification 24
 see also Algonquin languages
 Abnormalities, North American native language
 variation 756
 Aboriginal and Torres Strait Islander
 languages
 language relationships 80
 linguistic characteristics 80
 use of, Australia 79
 Aboriginal English 82
 Aboriginal languages
 use of 79
 see also Australian languages; Native American
 languages
 Absolute
 Afroasiatic languages 13
 Arrente 74
 Australian languages 81
 Caucasian languages 195
 South Asia 62–63
 Abun
 classification 1176
 nominal complex, genders 1177
 tone 1177
 verbal complex, Tense-Mood-Aspect 1176
 see also West Papuan languages
 Abure 631
 see also Tano languages
 Acehnese
 use of 99
 see also Austronesian languages
 Achagua
 stress 60
 see also Arawak languages
 Achi' 705–706
 speaker numbers 706*t*
 see also Mayan languages
 Achuan languages 504
 classification 505
 see also Hokan languages
 Acrolect 868
 Actionality (Aktionart), Russian verbs 907
 Active, Old Irish 453
 Active participles, in introflecting language 51,
 51*t*
 Acute, French orthography 428
 Adam, L., Cariban language classification 183
 Adamawa-Ubangi languages 771
 classification 768–769
 grammar 3
 noun classes 3
 phonetics/phonology 3
 study of 2
 SVO 3
 syntax 3
 use of 771
 Nigeria 771
 speaker numbers 2
 verbs 3
 workers in 2
 see also Benue-Congo languages; Dogon; Gur
 (Voltaic) languages; Kordofanian
 languages; Niger-Congo languages
 Adele 631
 see also Togo Mountain languages
 Ad Hoc Expert Group on Endangered Languages
 322
 dwindling domains 323–324
 language vitality assessment 324–325
 Adi 968–969
 see also Tani languages
 Adjectives
 Arabic 51
 Arikém 1106
 Balkan linguistic area 126, 127*t*
 Bantu languages 140
 Bengali 149
 Brahui 163
 Cubeo 1099
 Czech 277
 Danish 280
 Desano 1098, 1099
 Domari 296
 Dravidian languages 298
 English, Modern 330–331
 Finnish (Suomi) 414
 French 429
 Gondi 456
 Greek, Ancient 463
 Hebrew, Israeli 487
 Hokan languages 508
 introflecting language 51
 Kapampangan 579
 Kinyarwanda 608
 Kurukh 627
 Latin 642
 Luxembourgish 659–660
 Macuna 1098
 Mondé 1106
 Ossetic 815
 Persian, Old 853
 Portuguese 884
 Punjabi 887–888
 Ramaráma 1106
 Retuara/Tanimuca 1099
 Romani 899
 Secoya 1099
 Siona 1099
 Siriano 1099
 Slovak 978–979
 Slovene 983
 Somali 988
 Tariana 1051
 Telugu 1057
 Tiwi 1066–1067
 Tucano 1099
 Tucanoan languages 1098, 1099, 1099*t*
 Tungusic languages 1104
 Tupian languages 1106
 Tuyuca 1099, 1099*t*
 Wambaya 1162–1163
 Wolaitta 1181
 Adjukru 631
 see also Kwa languages
 Admiralties languages 99
 see also Austronesian languages
 Adpositionals, Oto-Mangean languages 822
 Adstratum relationship *see* Linguistic areas
 Afar (Qafar)
 number of speakers 272–273
 verb person marking 274–275, 275*f*
 see also Cushitic languages
 Affixation
 in introflecting language 52
 types
 circumfix 287
 infix 287
 prefix *see* Prefixes
 suffix *see* Suffix(es)
 see also Affixes; Clitic(s)
 Affixes
 diachronicity 287
 in isolating language 221
 nominal forms 290
 roots *vs.*, agglutinating *vs.* fusional languages 554
 verbs
 language types 288
 position determination 290
 Afghanistan
 languages
 Balochi 134, 538
 Brahui 162–163
 Dardic languages 282
 Indo-Iranian languages 531
 Iranian languages 537

- Afghanistan (*continued*)
 Kazakh 588
 Modern Persian 538, 850
 Uzbek 1145
 official languages, Pashto 538, 845
- Africa
 Islam *see* Islam
 languages *see* African languages
 as linguistic area 3–7
 consonants 4–5
 early work 4
 isopleth mapping 6, 6*f*
 logophoric marking 5
 ‘Pan-African properties,’ 4, 5*t*
 phonology 4
 quantitative evidence 5
 types 5*t*
 long-range comparisons 652–653
see also Areal Linguistics; Balkan linguistic area; Bantu languages; Chadic languages; Ethiopia; Ethiopian linguistic area (ELA); Europe, as Linguistic Area; Hausa; Highland East Cushitic (HEC) languages
- African-American English (AAE) 334
 USA 1125
see also African-American Vernacular English (AAVE)
- African-American Vernacular English (AAVE) 334–339
 aspectual markers 336
 auxiliary system 335–336
 discourse 335
 uncensored speech 335
 future 335–336
 habitual 336
 history/development 334–335, 337
 ‘Anglicist’ theory 337
 Creole theory 337
 influence from other languages 337
 variation theory 337
 lexicon 335
 misrepresentation 335
 negative concord 336
 negative inversion 336
 phonology 336
 initial voiced stop deletion 336
 plurals 336
 possessives 336
 pronouns 336
 resultative 336
 Southern White Vernacular English (SWVE) 335
 stigma 335
 syntax 335
 use of 334–335
see also African-American English (AAE); Creoles; English; Gullah; Pidgins
- African languages
 lexicostatistics 248–249
 SVO 5
see also specific languages
- Afrihili 76
- Afrikaans 7–12
 apartheid 8
 classification 251–252
 concord 9
 formal features 8
 Dutch *vs.* 8, 9
 history 7
 Bible translation 7–8
 influence from other languages 9, 11
 Bantu 11
 Creole Portuguese 9, 11
 Dutch 7–8, 310
 English 8
 Khoekhoe 7–8, 9, 11
 Malay 8, 9, 11
 influence on other languages, Fanagalo 411
 morphology 8–9
 as official language, South Africa 7
 script, Arabic 10
 SVO 9
 Taalmonument 9, 9*f*
 use of 7
 Namibia 7
- varieties 8
 Kaae Afrikaans 8, 9*f*
 Oosgrens Afrikaans 8, 9*f*
 Oranjeriver Afrikaans 8, 9*f*
 word order asymmetry 9
see also Dutch; Germanic languages; Indo-European languages; Krio; Zulu
- Afroasiatic languages 12–15, 206, 929
 absolutive 13
 classification 12, 250
 ergative 14
 geographical origin 12
 grammar
 nominal forms 13
 plural formation 13
 pronouns 13
 subject agreement 14
 investigational history 12
 Hamitic theory 12
 racial prejudice 12
 Nilo-Saharan languages *vs.* 773–774
 Nostratic theory 653–654, 786
 phonetics 14
 shared features 13
 use of 12
see also Akkadian; Amharic; Arabic; Berber languages; Chadic languages; Coptic; Cushitic languages; Eblaite; Egyptian; Ethiopian Semitic languages; Ge’ez; Hausa; Hebrew; Hebrew, Israeli; Hebrew, Pre-Modern; Highland East Cushitic (HEC) languages; Maltese; Nilo-Saharan languages; Omotic languages; Oromo; Semitic languages; Somali; Tigrinya; Wolaitta
- Agar *see* Dinka
- Agariya 736
see also Munda languages
- Agaw
 use of 272–273
see also Cushitic languages
- Agglutinating languages 291, 731, 732
 Balinese 117–118
 Cupeño 270
 Finnish 415–420
 fusional languages *vs.* *see* Fusional languages
 Georgian, Old 291
 Hurrian 516
 index of fusion 291
 index of synthesis 291
 Luganda 657–658
 Manambu 693
 Nenets (Yurak) 762
 other types *vs.* 732*t*
 Quechua languages 892
- Ahanta 631
see also Tano languages
- Ahmaogak, Roy, Inupiaq writing 535–536
- Aht *see* Nuuchahnulth
- Ainu 15–17
 adverbs 16–17
 applicative extension 17
 classification 249
 genetic affiliations 15–16
 nouns 16
 oral literature 15
 phonology 16
 assimilatory/dissimilatory processes 16
 consonants 16
 pitch accent system 16
 vowels 16
 plural verbs 16
 possession 16
 postpositions 17
 related languages, Japanese 557
 SOV 15–16
 subordinating conjugations 17
 suffixes 16
 use of, Japan 15
 verbs 16
 word order 17
- Aizi 624
see also Kru languages
- Aja (Aja-gbe) 631–632
see also Gbe languages
- Aka
 speaker numbers 772–773
see also Nilo-Saharan languages
- Akan 17–20
 consonants 18
 dialects 17
 Asante 17
 Fante 17
 dictionaries 17–18
 ethnography 18
 grammars (books) 17–18
 history/development 17
 influence on other languages 18
 morphology 19
 nouns 19, 632
 orthography 18
 phonology 18
 possessives 19
 postpositions 19
 serial constructions 19
 sociolinguistics 18
 SVO 19
 syntax 19
 tone 19, 632
 use of
 Ghana 17
 speaker numbers 18
 verbs 19
 vowel harmony 19
 vowels 19
 word order 19
 workers in 17–18
see also Kwa languages
- Akateko 705–706
 official recognition 705–706
 speaker numbers 706*t*
see also Mayan languages
- Akita 517
- Akkadian 20–22, 930
 dialects 20
 see also Assyrian; Babylonian
 dictionaries 21
 grammars (books) 21
 use of 20
 VSO 21
see also Afroasiatic languages; Assyrian; Babylonian; Eblaite; Persian, Old; Semitic languages; Sumerian
- Akkala Saami
 speaker numbers 911
see also Saami
- Aktionsart (actionality), Russian verbs 907
- Akuntsu
 classification 1106*t*
see also Tupian languages
- Akupem dialect 17
- Akuriyo
 use of 185*f*
see also Cariban languages
- Akyem dialect 17
- Alabama 738–739
 agreement type 741
 vowel length 740
- Alabama-Koasati 749
see also Muskogean languages
- Alacaluf 41
see also Andean languages
- Albania, Republic of
- Albanian 22
 Macedonian 663
 Romanian 901
- Albanian 22–24
 classification 251
 codification 23
 dialects 23
 Arbëresh 23
 Arvanitika 23
 Gheg 23
 Tosk 23
 geographic spread 22
 as official language 22
 origins/development 22

- phonemes 22
 scripts 23
 use of 22
 emigration effects 23
 Italy 545
 linguistic pockets 23
 vocabulary 22
 see also Balkan linguistic area; Indo-European languages
 Alesea-Siuslaw 750
 Aleut 373
 agreement system 373
 classification 251
 consonants 373
 dialects 373
 history 371
 influences from other languages, Russian 373
 labial stops 373
 pronouns 373
 use of 373
 see also Eskimo-Aleut languages
Algemene Nederlandse Spraakkunst 308
 Algeria
 Berber 152
 Songai languages 990–991
 Algin
 classification 25
 see also Algonquin languages
 Algonquin languages 24–30, 748
 circumfixes 287–288
 classification 24, 252
 ‘Central’ languages 24
 Eastern Algonquin 24
 Great Plains 25
 Illinois 25
 Indiana 25
 Michigan 25
 classifiers 28
 conjunct order 27
 demography 26
 derivational morphology 27
 dialects 754
 dictionaries 29
 documentation 28
 grammars (books) 29
 imperative order 27
 influences from other languages, English 24
 intransitive verbs 27
 mixed languages 28
 pidgins 28
 morphology 27
 nominals 27
 noun phrases 28
 nouns 27
 philology 28
 phonology 26
 possession 27
 syntax 28
 verb inflections 27
 vocatives 27
 word order 27, 28
 see also Abnaki; Algin; Arapaho; Central Siberian Yupik; Cree; Michif; Mobilian Jargon (Mobilier); Native American languages; Polysynthetic languages; Ritwan languages
 Algonquin-Ritwan hypothesis 651
 Algonquin-Wakashan languages 747–748
 Alienability
 Europe 394–395
 Somali 988
 Alladian 631
 see also Kwa languages
 Allomorphs, in agglutinating languages 417, 419
 Alphabets, Italian 547
 Alesea-Siuslaw *see* Penutian languages
 Altaic languages 30–33
 classification 250
 Nostratic theory 249
 influence on other languages
 Japanese 557
 Sino-Tibetan languages 970
 as ‘Micro-Altaic,’ 30
 Nostratic theory 653–654, 786
 Turkic-Mongol-Tungusic relationship 30
 as ‘Ural-Altaic,’ 30
 workers in
 Castrén, M A 30
 Polivanov, E D 31–32
 Ramstedt, Gustaf John 31
 see also Azerbaijani; Chuvash; Evenki; Japanese; Kazakh; Kirghiz; Korean; Mongolia; Mongol languages; Ryukyuan; Tungusic languages; Turkic languages; Turkish; Türkmen; Uralic languages; Uyghur; Uzbek; Yakut
 Alternation, in agglutinating languages 418–419
 Alutor
 classification 239
 speaker numbers 239–240
 see also Chukotko-Kamchatkan languages
 Alyawarra (Alyawarr)
 pronouns 90–91
 see also Australian languages
 Alyutor *see* Alutor
 American English
 African-American Vernacular *see* African-American Vernacular English
 British English *vs.*
 lexis 330
 morphology/syntax 332
 orthography 328
 phonology 329
 concord 336
 development 344
 use of 1123
 American languages, native *see* North American native languages
 American Sign Language (ASL) 956
 use of 1127
 Amerindian languages, long-range comparisons 649, 652, 653
 Amharic 33–36
 accent 34
 case system 34–35
 consonants 34, 34*t*
 converbs 35
 earliest records 33
 influence from other languages, Cushitic languages 33
 IPA *vs.* 34
 morphology 34
 negation 35
 phonology 34
 plurals 34–35
 pronoun object markers 35
 SOV 36
 syllable structure 34
 syntax 36
 TMA marking 35
 use of 33, 382–383, 929
 Ethiopia 33
 as first language 33
 as L2 33
 verbs 35
 vowels 34
 word order 36
 writing 33–34
 see also Afroasiatic languages; Ethiopian linguistic area (ELA); Ethiopian Semitic languages; Ge’ez; Semitic languages
 Amis
 classification 421
 dialects 421
 research history 423
 speaker numbers 421
 see also Formosan languages
 Ammonite, Phoenician *vs.* 854
 Amto-Musian languages, geographical distribution 840–841
 Amuesha 41
 predicate structure 60
 see also Andean languages
 Amusgo
 classification 819–821
 syllable onsets 821–822
 see also Oto-Mangean languages
 Amuzgoan languages 751
 see also Oto-Mangean languages
 Amwi 595
 Anal
 classification 968–969
 see also Kuki-Chin languages
 Analytic case relations, Balkan linguistic area 125
 Analytic gradations, adjectives, Balkan linguistic area 126, 127*t*
 Analytic subjunctives, Balkan linguistic area 127, 128*t*
 Anatolian languages 36–38
 classification, genetic classification 246
 dialects 37
 noun morphology 37
 historical aspects 36
 iterative 37
 lexicon 37
 morphology 37
 nouns 37
 origins 38
 particles 37
 phonology 36
 vowel system 36–37
 reconstruction 246
 verbs 37
 see also Indo-European languages
 Ancestral languages, genetic classification 246
 Ancient Egyptian *see* Egyptian
 Ancient Greek *see* Greek, Ancient
 Andaqi, long-range comparison 653
 Andean languages 40–42
 classification 40
 definition 40
 ergative 40
 extinct varieties 41
 types 40
 use of
 Argentina 41
 Bolivia 41
 Chile 41
 Colombia 40
 Ecuador 41
 Panama 40
 Patagonia 41
 Peru 41
 Venezuela 40
 see also Alacaluf; Amuesha; Arawak languages; Aymara; Cariban languages; Chibchan languages; Quechua languages
 Andersen, Torben, Dinka 293
 Andi
 vowels 193, 193*t*
 see also Caucasian languages
 Andorra, Catalan 188, 191
 Anem
 geographical distribution 841
 see also West New Britain languages
 Angal Enen *see* Mendi (Angal Enen)
 Angan languages
 classification 1087
 see also Trans New Guinea languages
 Angkuic languages 727
 see also Palaung-Wa languages
 ‘Anglicist’ theory, African-American Vernacular English development 337
 Anglo-Saxon *Chronicle* 357
 Angola, Portuguese 883
 Angry register, Bikol 160
 Anhui
 classification 969
 see also Hui languages
 Animere 631
 see also Togo Mountain languages
 Anticausative-prominence, Standard Average European (SAE) languages 393–394
 Antilles 307
 Antonyms/antonymy *see* Negation
 Anufo 631
 see also Tano languages
 Anyi 631
 see also Tano languages
 Apalachee 739

- Apalai
 phonology 183–184
 reduplication 184
 Apartheid, Afrikaans 8
 Apatani
 classification 968–969
 see also Tani languages
 Apinajé
 as ergative language 668
 word order 667–668
 see also Macro-Jê languages
a posteriori languages, artificial
 languages 77
a priori languages, artificial languages 77
 Arabic 42–50
 accusative 46
 adjectives 48
 agreement 48
 Classical *see* Arabic, Classical
 classification 250
 derivational morphology 45
 dialects 43, 54
 agreement 49
 auxiliary verbs (aspect neutralizers) 49
 Bedouin 54
 case system 47
 ‘genitive exponents,’ 49
 modern Arabic dialect groups 54
 morphology 47
 phonology 45
 pronouns 47
 Sedentary 54
 syntax 49
 types 43–44
 verbs 47
 word order 49
 diminutives 52
 equational sentences 48
 genders 46
 genitive 46
 history of 42
 oral traditions 42–43
 imperfective 47
 imperfect tense 46–47
 inflectional morphology 52
 affixations 52
 dual 52
 pronominal subject markers 52, 52*t*
 sound plural 52
 influence on other languages 43
 Bengali 148
 Berber 153
 Domari 295, 296
 French 429
 Hindi 495–496
 Kashmiri 582–583
 Malayalam 680–681
 New Iranian languages 538
 Punjabi 889
 Spanish 1020
 Yanito 1202
 as introflecting language 50–53
 language spread 318
 Middle 932
 modern *see* Modern Standard Arabic
 Modern Southern 931
 morphology 45
 negation 48
 neologisms 45
 nominal annexations 48
 nominative 46
 North *see* Arabic, North; *below*
 nouns 51
 adjectives 51
 broken plurals 52
 ‘broken’ plurals 52
 comparatives 52
 diminutives 52
 elatives 52
 finite verb stems 51, 51*t*
 noun inflexion 46
 singular nouns 51
 ‘sound’ plurals 52
 superlatives 52
 number system 46
 cardinal numbers 48
 as official language 42
 Israel 42, 485
 Mauritania 42
 Oman 42
 Old Southern 931
 OVS 49
 perfect 46, 51
 phonology 44
 consonants 44, 44*t*
 religious use 44
 syllabic structure 44
 vowels 44
 pronouns 46, 47*t*
 relative clauses 48
 root and pattern 45–46, 50, 50*f*
 consonantal roots 50
 noun stems 50
 verb stems 50
 Southern *see below*
 subordination 48
 SVO 47
 syntax 47
 tense and aspect 47
 tenses 45, 46*t*
 see also specific tenses
 use of 42, 929
 Chad 42
 Iran 42
 Islam 42
 Nigeria 42
 Turkey 42
 variation *see* Arabic, variation
 verbs 50, 51*t*
 active participles 51, 51*t*
 passive participles 51, 51*t*
 verb inflexion 46
 vocalic melody 51
 VSO 47
 word order 47
 see also Afroasiatic languages; Arabic; Arabic,
 as introflecting language; Arabic,
 variation; Aramaic; Central Siberian
 Yupik; Modern Standard Arabic (MSA);
 Morphological Types; Persian, Modern;
 Polysynthetic languages; Punjabi; Semitic
 languages; Syriac; Turkic languages;
 Turkish; Urdu
 Arabic, Classical 43, 932
 influence on other languages 43
 see also Islam; Semitic languages
 Arabic, Middle 932
 see also Semitic languages
 Arabic, North 931
 Hasaitic 931–932
 Hismaic 932
 Oasis dialects 932
 Safaitic 932
 Thamudic 932
 see also Semitic languages
 Arabic, Southern
 Modern 931
 Old 931
 see also Semitic languages
 Arabic, variation 53–56
 Arab world countries 53
 common variations 54
 dialect contact 55–56
 education 55
 gender 55
 social class 55
 historical aspects 53
 British and French influence 53–54
 Ottoman Turkish Empire 53–54
 Standard Arabic 54
 see also Arabic; Berber languages
 Arabic Persian, influence on other languages,
 Azerbaijani 112
 Arabic script
 Afrikaans 10
 Azerbaijani 111
 Fulfulde 430
 Kazakh 589
 Kirghiz 611
 Malagasy 674–675
 Modern Persian 850
 Pashto 846
 Turkmen 1117
 Uyghur 1143
 Uzbek 1146
 Aramaic 56–59, 932, 934
 classification 250
 dialects 57, 929
 spoken dialects 58
 Syriac *see* Syriac
 influence from other languages, Judeo-Arabic 568
 influence on other languages 57
 Jewish languages 566
 Yiddish 567, 1205
 Late 934
 literary dialects 56
 Middle 57–58, 934
 Modern (Neo-Aramaic) 934
 use of 934
 see also Semitic languages
 Official (Imperial) 57, 934
 see also Semitic languages
 Old 57, 934
 writings 57–58
 see also Semitic languages
 origin/expansion 56
 religious communities 57
 use of 929
 Azerbaijan 58
 biblical texts 483
 East Syrian Christians 58
 Egypt 56
 geographical distribution 56
 Iran 56
 Judaic commentaries 58
 Kurdistan 58
 Syrian Orthodox Church 58
 in Talmud 483
 Turkey 58
 see also Afroasiatic languages; Arabic; Hebrew;
 Iranian languages; Modern Standard
 Arabic (MSA); Persian, Modern; Semitic
 Languages; Sogdian; Syriac
 Aranama 751
 see also Native American languages
 Arapaho
 speaker numbers 26
 stress 26
 see also Algonquin languages
 Arapesh (Bukiyip: Muhiang)
 class systems 1078
 see also Torricelli languages
 Arara
 use of 185*f*
 see also Cariban languages
 Araucanian 752
 see also Native American languages
 Arawakan 750
 see also Native American languages
 Arawak languages 40
 affiliations, Mapudungan languages 701
 classification 252–253
 classifiers 60, 61
 as endangered languages 59
 genders 61
 genetic unity 59–60
 influence on other languages 59
 lexicon 61
 negation 61
 nouns 61
 plurals 61
 predicate structure 60
 prefixes 60
 pronominal suffix loss 60
 stress 60
 suffixes 60
 tones 60
 use of 59
 verbs 60
 workers in 60
 see also Achagua; Andean languages;
 Guarekena (Warekena)

- Arbëresh 23
 Archeology, Indo-European language classification 530
 Archi
 phonetics 193, 193*t*
 see also Caucasian languages
 Ardabil
 classification 112–113
 see also Azerbaijanian
 Areal linguistics 62–68
 definition 62
 genetic relationships 66
 language subgroupings 65, 65*t*
 linguistic reconstruction 65
 Native American languages 746
 see also Africa; Africa, as linguistic area;
 Balkan linguistic area; Ethiopia;
 Ethiopian linguistic area (ELA);
 Kashmiri; Linguistic areas;
 Southeast Asian languages;
 Wakashan languages
 Arem 728–729
 see also Chut languages
 Argentina
 Andean languages 41
 Guaraní 467
 Ingáin 666–667
 Italian 545
 Macro-Jê languages 666–667
 Mapudungan languages 701
 Quechua 891
 Argobba
 use of 382–383
 see also Ethiopian Semitic languages
 Argumentatives *see* Diminutives
 Ari 805
 long-range comparisons 652–653
 see also Omotic languages
 Arikara 749
 see also Caddoan languages
 Arikém
 adjectives 1106
 classification 1106*t*
 tone system 1106
 see also Tupian languages
 Armenia 68
 Armenian languages 68–72
 alphabet 70*t*
 classification 251
 genetic classification 246
 development 69–70
 examples 70
 Greek *vs.* 68–69
 Hübschmann, Heinrich 68–69
 Indo-Iranian *vs.* 68–69
 nouns 69
 phonology 69
 pronouns 69
 reconstruction 246
 sound correspondences 650–651
 subordinate clauses 69
 use of 68
 verbal conjugations 69
 vocabulary 69
 written records 69
 see also Indo-European languages;
 Romani
 Aromanian 901
 Arrernte 72–75
 absolute 74
 Bible translation 73
 changes 73–74
 classification 72–73
 consonants 73, 73*t*
 ergative 74
 history 72–73
 kinship interactions 74
 monosyllabic words 74
 morphology 74
 pronouns 74
 study of 73
 syllables 73–74
 use of 72–73
 geographical distribution 72–73, 72*f*
 vowels 73–74
 see also Australian languages; Kaytetye;
 Morrobalama; Warlpiri
 Ars Magna (Raymundus Lullus) 76
 Arte de la Lengua Bisaya de la Provincia de Leite 915
 Articles, Standard Average European (SAE)
 languages 393–394
 Artificial languages
 a posteriori languages 77
 a priori languages 77
 auxiliary languages 75–76, 77
 classification systems 77
 constructors 75
 definition 75
 grammar 77
 hypothesis testing 76
 idioms 77
 International Auxiliary Language Association 77
 Sapir-Whorf hypothesis 76
 semantics 77
 vocabulary 77
 workers in
 Brown, James Cooke 76
 de Wahl, Edgar 77
 Hildegard of Bingen, Saint 76
 Llull, Ramón 76
 Peano, Giuseppe 77
 Schleyer, Johan Martin 76
 Sudre, Francois 76
 Wilkins, John 76
 see also Esperanto
 Arúa
 classification 1106*t*
 see also Tupian languages
 Aruba, Dutch 307
 Arvanitika, Albanian dialects 23
 Asante dialect, Akan 17
 Ashkun 787
 dialects 787
 see also Nuristani languages
 A sho Chin 968–969
 see also Kuki-Chin languages
 Asia
 South *see* South Asian languages
 Southeast *see* Southeast Asian languages
 Asian languages 94–95
 Malaysia 94–95
 Thailand 94–95
 see also Austroasiatic languages
 Asmat-Kamoro languages
 classification 1087
 see also Trans New Guinea languages
 Asoka
 origins/development 523
 see also Indo-Aryan languages
 Aspectual character 907
 Aspectual class 907
 Aspectual marking, sign language morphology 952
 Aspirated voiceless stops, nonnative English 360
 Aspiration
 Dardic languages 283
 Scots Gaelic 927
 Assamese 78–79
 Bangladesh 78
 Bengali *vs.* 78
 classification 522
 converbs 995
 dialects 78
 Hindi *vs.* 78
 morphology 78
 number of speakers 523
 Oriya *vs.* 78
 phonetics 78
 phonology 525–526
 vowels 526–527
 pidgin 78
 syntax 78
 written 78
 see also Indo-Aryan languages
 Assibilation, in agglutinating languages 418–419
 Assyrian 20
 influence from other languages, Phoenician 854
 ‘Assyrians,’ 1033
 Astori
 classification 282
 see also Astor languages
 Astor languages
 classification 282
 see also Astori; Shina languages
 Asuri 736
 see also Munda languages
 Ata
 use of 841
 see also West New Britain languages
 Atacameño 41
 Atakapa 749
 see also Muskogean languages
 Atayalic languages
 classification 421
 see also Formosan languages
 dialects 421
 dictionaries 423
 research history 422–423
 Athabaskan–Eyak–Tlingit (AEC) 743–745
 classification 252
 see also Na-Dene languages
 Atlantic Congo languages 770
 classification 768–769
 subgroups 770
 use of 770
 see also Niger-Congo languages
 Attié 631
 see also Kwa languages
 Augmentative
 Bantu languages 140
 Creek 264
 Crow 269
 French 428–429
 North American native languages 757
 Nuuchahnulth (Nootka) 789
 Tupian languages 1106
 Xhosa 1188
 Australia 79–84
 languages *see* Australian languages; *specific languages*
 Australian English 79, 82
 loanwords 82
 regional variation 82
 social variation 82
 use of 79, 82
 Australian languages 84–92
 Aboriginal 79
 Aboriginal English 82
 absolute 81
 avoidance language 91
 classification 84, 250
 Blake 84
 Capell 84
 O’Grady 84
 community languages 81
 consonants 86*f*
 Dutch 307
 ergative 81, 87, 88–89
 Fijian 412
 Finnish 413
 geographical distribution 84, 85*f*
 Gujarati 468
 Italian 545
 Kala Lagaw Ya 79
 lexical roots 84
 Meryam Mer 79
 Modern Greek 464
 morphology 87
 Morrobalama 735
 non-Pama-Nyungan languages 89
 noun classes 89
 noun phrases 89
 phonology 86
 pidgins and Creoles 81
 pronominal forms 90–91, 91*t*
 pronouns 89–90, 90–91
 secret languages 91
 ‘mother-in-law’ languages 91
 semantics 90
 sign languages 91
 SOV 90

- Australian languages (*continued*)
 stop sounds 86
 syllables 86
 syntax 87
 Torres Strait Islander 79
 vowels 86
 word order 90
 workers in
 Blake 84
 Capell 84
 Dixon, R M W 250
 O'Grady 84
see also Aboriginal languages; Alyawarra (Alyawarr); Arrernte; Austronesian languages; Dien (Diyari); Gamilaraay; Guugu Yimithirr; Hungarian; Jiwarli; Kalkutungu; Kayardild; Kaytetye; Morrobalama; Pitjantjatjara; Tiwi; Wambaya; Warlpiri
- Australian Sign Language (Auslan) 82
- Austria
 German 444
 Hungarian 514
 Slovene 981
- Austic hypothesis 92–94
 morphology 92
 Nicobarese languages 92
 phonology 92
 Schmidt, Wilhelm 92
 syntax 92
see also Austroasiatic languages; Austronesian languages; Mon-Khmer languages; Sino-Tibetan languages
- Austroasiatic languages 94–96
 classification 250
 development *see* Austric hypothesis
 influence on other languages, Sino-Tibetan languages 970
 morphology 95
see also Asian languages; Austric hypothesis; Austronesian languages; Burushaski; Khasi languages; Mon; Mon-Khmer languages; Munda languages; Santali; Sino-Tibetan languages; Southeast Asian languages; Wa
- Austronesian languages 96–105
 classification 250, 685, 685*f*
 clause structure 100–101
 comparative reconstruction 99
 consonants 100
 development 687
see also Austric hypothesis
 external genetic relationships 102
 geographical spread 97, 103*f*
 Samoa 102
 Tonga 102
 historical interpretation 102
 population movements 102
 historical studies 98
 internal genetic relationships 98
 Japanese development 557
 morphology 100
 number of speakers 96–97
 phonology 100
 possessive forms 101
 possessives 101
 religious influences 97
 structural diversity 98
 subgroups 99
 Admiralties subgroup 99
 Central and Eastern Oceanic subgroup 99–100
 Oceanic subgroup 99
 Western subgroup 99
 Tai-Kadai relation *see* Austro-Tai hypothesis
 use of 96, 97
 Brunei 97
 Indonesia 97, 99
 Madagascar 97
 Malaysia 97, 99
 Philippines 97, 99
 Singapore 97
 Sulawesi 99
 Sumbawa 99
 Taiwan 97, 105
- workers in
 Codrington, R H 98
 Dempwolff, Otto 98
 Dyren, Isidore 98
 Panduro, Lorenzo Hervas 98
 Reland, Hadrian 98
 von der Gabeltzen, H C 98
see also Acehnese; Admiralties languages; Australian languages; Austroasiatic languages; Austro-Tai hypothesis; Ayatalic languages; Benue-Congo languages; Bikol; Cebuano; Creoles; Fijian; Flores languages; Formosan languages; Hiligaynon; Ilocano; Japanese; Javanese; Madurese; Malagasy; Malay; Malayo-Polynesian languages; Malukan languages; North Philippine languages; Papuan languages; Pidgins; Riau Indonesian; Samar-Leyte; South Asian languages; Tagalog; Tamambo; Trans New Guinea languages; Vurës; West Papuan languages
- Austro-Tai hypothesis 105–107
 Benedict, P K 105
 Ostapirat, W 105
 Sagart, L 106
see also Austronesian languages; Tai Languages
- Austro-Tai languages, Benedict, Paul 249
- Autolexical theory, West Greenlandic 1173–1175
- Auvergnat *see* Occitan
- Auxiliary languages, artificial languages 75–76, 77
- Avair
 morphology 194, 194*t*
 phonetics, vowels 193, 193*t*
see also Caucasian languages
- Avestan 107–108, 537
 classification 251–252
 inflectional morphology 108
 lexicon 108
 manuscripts 107–108
 nominal systems 539
 nouns 533
 Old Persian divergence 107
 Old *vs.* Younger 107
 oral tradition 107
 origin/development 107
 pronominal systems 539
 Sanskrit *vs.* 918–919
 texts 107
 verbs 108, 539
 word order 534
 Zoroastrianism 107
see also Indo-European languages; Indo-Iranian languages; Iranian languages; Pashto; Persian, Modern; Persian, Old; Sanskrit; Sogdian
- Avikam 631
see also Kwa languages
- Awá
 classification 224
 use of 224
- Awakateko 705–706
 speaker numbers 706*t*
see also Mayan languages
- Aweti
 classification 1106*t*
 morphology, ideophones 1106–1107
see also Tupian languages
- Awutu 631
see also Guang languages
- Awyu-Dumut languages
 classification 1087
see also Trans New Guinea languages
- Ayapa Zorqué *see* Mixe-Zoquean languages
- Ayatalic languages
 classification 250–251
see also Austronesian languages; Formosan languages
- Aymara 752
 agglutinating structure 109
 Cauqui languages *vs.* 108
 dialects 109
 dictionaries 109
 evidentiality 109
 grammar 109
 history 109
 Jaqaru languages *vs.* 108
 lexicon 108–109
 morphology 109
 nominalization 109–110
 nouns 109
 Quecha languages *vs.* 108–109
 suffixes 109
 use of 108
 velar nasal consonants 109
 verbs 109
 vowels 109
 workers in, Bertonio, Ludovico 109
see also Andean languages; Native American languages; Quechua languages
- Aymaran languages 41
 Quechua *vs.* 891
 use of 41
- Ayrum
 classification 112
see also Azerbaijani
- Ayurú
 classification 1106*t*
see also Tupian languages
- Azerbaijan
 Aramaic 58
 Armenian 68
 Azerbaijani 110, 1112
 Caucasian languages 192
 Georgian 442
 Azerbaijani 110–113, 1109, 1112
 as agglutinative language 111
 dative forms 112
 dialects 112
 grammar 112
 influence from other languages
 Arabic-Persian 112
 Persian 112
 Russian 110–111
 Turkish 110–111
 language contacts 110
 lexicon 112
 origin/history 110
 perfect 112
 perfect markers 112
 phonology 111
 present-tense maker 112
 related languages 110
 sound harmony 111
 use of 110
 Azerbaijan 110, 1112
 Iran 112–113, 1112
 speaker numbers 110
 vowel harmony 111
 vowels 111
 written language 111
see also Altaic languages; Ardabil; Ayrum; Turkic languages; Turkish; Türkmen
- Aztec *see* Nahuatl
- Aztecan
 classification 1139
see also Uto-Aztecan languages
- Aztec-Tanoan languages, historical aspects 747–748
- ## B
- Babylonian 20, 930
 influence from other languages, Phoenician 854
see also Akkadian; Semitic languages; Sumerian
- Baby talk
 North American native language variation 757
 vocabulary, Hopi 513–514
- Bactrian 115–116, 538
 classification 251–252
 declensions 540
 definite articles 540
 ergative 115
 future 116
 genders 115, 540
 Greek script 115
 history 115

- past tenses 541
perfect 115
pronouns 541
verbs 115
see also Iranian languages
- Badaga, Malayalam *vs.* 682
- Bahnar
speaker numbers 726
see also Bahnaric languages
- Bahnaric languages 724
classification 725–726
dialects 726
morphology, verbs 725
use of 725–726
speaker numbers 726
see also Laven (Bolovent); Mon-Khmer languages
- Bai languages
classification 969
morphology 970
see also Sino-Tibetan languages
- Baining languages
geographical distribution 841
see also Papuan languages
- Bakairi
geographical distribution 185f
morphemes 184–185
phonology 183–184
see also Cariban languages
- Balangaw (Balango)
phonology 784
see also North Philippine languages
- Balango *see* Balangaw (Balango)
- Balanta 770
see also Atlantic Congo languages
- Bali 116
- Balinese 116–119
as agglutinating language 117–118
consonants 117
dialects 117
dictionaries 118
grammars (books) 118
history 116
literary tradition 116–117
influence from other languages, Javanese 116–117
morphosyntax 117
orthography 117
lontar writing 117
Old Javanese script 117
Roman script 117
phonology 117
sociolinguistics 116
status importance 116–117
syllable structure 117
use of 99, 116
vowels 117
word order 118
see also Austronesian languages; Javanese; Malayo-Polynesian languages
- Balkan linguistic area 62, 119–134
adjectives 126, 127t
analytic case relations 125
analytic subjunctives 127, 128t
causation 132
concord 125
consonants 122
derivational morphology 131
evidentiality 130–131
future 129, 129t
negated future tense 129t
in past as conditional 129, 129t
will/have future tense 128, 129t
genitive-dative merging 125
grammaticalized definiteness 123
have perfect tense 129
lexicon 131
morphosyntax 123
numeral formation 127, 127t
perfect 130t
phonology 122
possessives 124
postpositions 122
pronominal object doubling 124
prosody 123
reduplication 124
replication 124
resultative 126
resumptive clitic compounds 124
semantics 131
sociolinguistics 132
language prestige 132, 132f
stressed schwa 122
SVO 131
vowel raising 122
vowel reduction 122
word order 130
clitic ordering 130
constituent order 131
see also Africa; Africa, as linguistic area; Albanian; Areal linguistics; Ethiopian linguistic area (ELA); Europe, as linguistic area; Greek, ancient; Greek, modern; Latin; Macedonian; Old Church Slavonic; Romani; Romanian; Sanskrit; South Asian languages; Southeast Asian languages; Turkish
- Balkans
definition 119
languages of 119, 120
linguistic history 121
see also Balkan linguistic area; Europe, as Linguistic Area
- Balkan Sprachbund, Romanian 902
- Balochi 134–135
case system 134
consonants 134
dialects 134
lexicon 135
morphology, declensions 540
oral tradition 134
use of 134, 538
verbs 134–135
see also Iranian languages; Pashto
- Baltic languages
classification 251–252
see also Indo-European languages
- Balto-Slavic languages 135–136
Brugmann, K 135
definition 135
Endzelin, J 135–136
Meillet, A 135–136
phonology 135
possessives 136
see also Belorussian; Bulgarian; Church Slavonic; Czech; Latvian; Lithuanian; Macedonian; Old Church Slavonic; Polish; Russian; Slavic languages; Slovene; Sorbian
- Baluchi
classification 251–252
see also Iranian languages
- Banda languages 3
see also Adamawa-Ubangi languages
- Bangladesh
Assamese 78
Bengali 148
Burmese 170
Indo-Aryan languages 522
Khasi 595
Munda languages 736
Sino-Tibetan languages 968
Urdu 522–523, 1133
- Baniata *see* Touo (Baniata)
- Baniwa
classifiers 61
verbs 60
see also Arawak languages
- Bantu languages 136–143
adjectives 140
augmentative 140
classification 771–772
clicks 1017–1018
concord 140
consonants 139
Dahl's law 139
demography 136
diminutives 1018–1019
downstep 139
Efik *vs.* 314–315
endangered types 137
future 141
genders 140
influence on other languages
Afrikaans 11
Luo 659
Katupha's law 139
Meinhof's law 139
morphemes 140–141
morphology 140
multilingual communities 136–137
nouns 140
classes 140
phrases 141–142
obligatory contour principle (OCP) 139–140
origin/history 137
phonology 138
pronouns 140
relative markers 141
spirantization 139
SVO 141–142
syntax 141
tenses 141
tonality 139
tone spreading 139–140
types 138
verbs 140
vowels 138–139
length 139
word order 141–142
see also Africa; Africa, as linguistic area; Bantu languages, Southern; Benue-Congo languages; Fanagalo; Kinyarwanda; Luganda; Mambila; Niger-Congo languages; Nyanja; Shona languages; Swahili; Xhosa; Zulu
- Bantu languages, Southern 1017–1020
classification 1017
click consonants 1018
concord 1018
definition 1017
morphology 1018–1019
noun class 1018
perfect 1017
phonology 1018
prefixes 1018–1019
special characteristics 1018–1019
SVO 1019
syntax 1019
use of 1017
vowel systems 1018
word order 1019
see also Bantu languages; Shona languages; Xhosa; Zulu
- Bará *see* Waimaja/Bará
- Barasano/Taiwano
morphemes 1096
nasalization 1095–1096
speaker numbers 1092t
verbs 1099–1100
vowels 1092–1093
word order 1096
see also Tucanoan languages
- Barbacoan languages 40–41
see also Andean languages
- Bare
pronominal suffix loss 60
see also Arawak languages
- Bari *see* Chibchan languages
- Barrett, Samuel A, Pomo language classification 878
- Baru/Lavé
speaker numbers 726
see also Bahnaric languages
- Barupu (Warupu) 974
see also Skou languages
- Basay-Tobiawan
classification 421
see also Formosan languages
- Bashkarik
vowels 526–527
see also Indo-Aryan languages

- Bashkir 143–144
 consonants 143–144
 dialects 144
 origin/history 143
 phonology 143
 related languages 143
 SOV 143
 vowel harmony 143–144
 vowels 143
 written language 143
see also Kazakh; Tatar; Turkic languages
- Basic words, lexicostatistics 248
- Basilect 867
- Basque 144–147
 and Amerind 655
 articles 146
 classification 249
 consonants 145
 demonstratives 146
 dialects 145
 ergative 145–146
 grammar 145–146
 morphology 145
 noun phrases 146
 phonetics 145
 relation to other languages 144
 use of
 France 144–145
 historical areas 145
 Spain 144–145
 verbs 145–146
 vowels 145
 word order 146
 SOV 146
 SVO 146
see also Spanish
- Bassa languages 770
 Liberia 624
 speaker numbers 623
see also Kru languages
- Baule 631
see also Tano languages
- Bazaar Hindustani 499
- Bazaar Malay, Riau Indonesian *vs.* 895–896
- Bedawiye *see* Beja (Bedwari: Bedawiye)
- Bedouin dialect 54
- Beifang
 classification 214
 speaker numbers 214*t*
see also Mandarin
- Beijing Mandarin
 classification 214
 speaker numbers 214*t*
see also Mandarin
- Beja (Bedwari: Bedawiye)
 use of 272–273
see also Cushitic languages
- Beke, C T 13
- Belgium
 Dutch 307, 310
 French 427
 German 444
- Belize, Arawak languages 59
- Bella Coola 749
see also Salishan languages
- Belorussian 147–148
 classification 251–252, 974–975
 Cyrillic alphabet 147
 declensions 976–977
 influence on other languages,
 Yiddish 1205
 lexicon 147
 morphology 147
 nouns 147
 origin/development 147
 orthography 147
 phonology 147
 Russian *vs.* 147
 Ukrainian *vs.* 147, 1122–1123
 use of 147
 verbs 147
see also Balto-Slavic languages; Polish; Russian;
 Slavic languages; Ukrainian
- Benedict, Paul 105, 249
- Bengali 148–150
 adjectives 149
 aspirated *vs.* unaspirated sounds 148
 Assamese *vs.* 78
 classification 251–252, 522
 correlative 148
 dental *vs.* palatal sounds 148
 dialects 148
 habitual 149
 impersonal structures 149–150
 influence from other languages 148
 locative ending 149
 morphology 148
 Nepali *vs.* 764
 nouns 148
 genitive nouns 148–149
 number of speakers 523
 object case 149
 as official language 148
 onomatopoeia 150
 orthography 148
 Devanagari script 148
 passives 150
 perfect 149
 phonology 148, 525–526
 postpositions 149
 pronouns 149
 special features 150
 syntax 148
 tonal system 525–526, 526–527
 verbs
 compound verbs 996
 conjugation 149
 intransitive verbs 150
 nonfinite verb forms 149
 vowels 526–527
 word order 148
 SOV 148
 writing systems, Nagari 524
see also Hindi; Indo-Aryan languages; Indo-European languages; Indo-Iranian languages; Persian, Modern; Persian, Old; Sanskrit; South Asian languages
- Benin
 Gur 770
 Gur languages 472
 Kwa 771
 Kwa languages 630
 Mande 769–770
 Yoruba 1207
- Benue-Congo languages 771
 classification of 771*f*
 changes 151
 concord 151
 Greenberg, Joseph H 150
 morphology 151
 noun classes 151
 phonology 151–152
 subgroups 151
 use of 771
 geographical locations 151
 Nigeria 771
 verbs 151
 word order 151
 SVO 151
see also Adamawa-Ubangi languages;
 Austronesian languages; Efik; Niger-Congo languages
- Ben Yehuda, Eliezer, Israeli Hebrew development 485
- Beothuk 751
 classification 26
see also Algonquin languages; Native American languages
- Berber languages 12, 152–158
 adjectival schemes 153–154, 154*t*
 aspect 156, 156*t*
 aspectual inflections (Taqbaylit) 154, 154*t*
 case 154, 155*t*
 classification 250
 Nostratic theory 249
 clitics 155
 consonants 153
 constituent order 154
 dialects 154
 diminutives 154
 genders 154
 head marking 155
 imperfective 156
 influence from other languages, Arabic 153
 morphology 153
 negation 156–157, 157*t*
 negative form 156, 157*t*
 nominal schemes 153–154, 154*t*
 noun phrases 154
 perfect 157*t*
 personal affixes 155
 phonetics/phonology 153
 plurals 154, 155*t*
 predicate nominals 155
 attributive predication 156, 156*t*
 possession 156
 progressive 156*t*
 relative clauses 155, 155*t*
 resources 157
 resultative 156
 stem composition 153–154
 use of 152
 Algeria 152
 Egypt 152–153
 geographical distribution 153*f*
 Libya 152–153
 Mali 152
 Morocco 152
 Niger 152
 Tuaregs 152
 Tunisia 152–153
 verbal derivational 154, 154*t*
 word order 155, 155*t*
 VSO 154
see also Afroasiatic languages; Arabic, variation
- Berbice Dutch Creole, classification 249*t*
- Bernolák, Anton 980
- Berta 775*f*
- Bertonio, Ludovico 109
- Bete languages
 use of 623–624
 speaker numbers 623
see also Kru languages
- Betoi 41
see also Andean languages
- Bhadrawahi
 vowels 526–527
see also Indo-Aryan languages
- Bharatesvarababubalirasa 468
- Bhattani Punjabi
 classification 886
see also Punjabi
- Bhili
 phonology 525–526
see also Indo-Aryan languages
- Bhoi 596
- Bhumij (Mudari) 736
see also Munda languages
- Bhutan
 Indo-Aryan languages 522
 Nepali 764
- Biaspectual verbs, Slovak 979
- The Bible
 Aramaic 483
 Hebrew 482
 translations *see* The Bible, translations
see also Aramaic; Syriac
- The Bible, translations
 Afrikaans 7–8
 Arrernte 73
 Dutch 308
 Formosan languages 422
 Gamilaraay 438
 German 445
 Krio 618
 missionary movements *see* SIL (Summer Institute of Linguistics)
 Wa 1155
 West Greenlandic 1175
 Yoruba 1207
- Bickerton, Derek, Creoles 861
- Bikat Kahani 498

- Buin 841
 see also South Bougainville languages
 Bukharan *see* Judeo-Persian
 Bulgaria
 Gagauz 1112
 Macedonian 663
 Romani 898
 Romanian 901
 Turkish 1112
 Bulgarian 168–170
 classification 251–252, 974–975
 dialects 168
 imperfective 168
 influence on other languages, Macedonian 663
 morphology, declensions 976–977
 perfect 169
 phonology 168
 related languages 168
 resultative 169
 tenses 168
 future 169
 renarrative construction 169
 word order 169–170
 see also Balto-Slavic languages; Church Slavonic; Macedonian; Old Church Slavonic; Slavic languages; Slovene
- Bunjwali
 classification 282
 see also Kashmiri languages
 Bunun
 classification 421
 dialects 421
 research history 423
 see also Formosan languages
 Burak languages 3
 see also Adamawa-Ubangi languages
 Burji
 noun morphology 491*t*
 phonology 490–491, 490*t*
 use of 488–489, 488*t*
 see also Highland East Cushitic (HEC) languages
 Burkina Faso
 Gur 770
 Gur languages 472
 Mande 769–770
 Songai languages 990–991
 Burma/Myanmar
 Burmese 170
 Hindi 495
 Karen languages 581
 Khmuic languages 727
 Mon 718, 727
 Mon-Khmer languages 725
 Palaung-Wa languages 727
 Pali 830
 Sino-Tibetan languages 968
 Tai-Kadai languages 105
 Tai languages 1039
 Tibetan 1060–1061
 Wa 1155
 Waic languages 728
 Burmese 170–175
 classification 968–969
 compounding 173
 consonants 171–172
 derivational morphology 173
 forms of address 173
 glottal stops 172
 history 170
 influence on other languages, Wa 1156
 literacy 173
 morphemes 172–173
 morphology 173
 noun case markers 173
 phonetics/phonology 171
 postpositions 173
 pronouns 173
 script 170
 affricates 171
 alphabet 170
 consonants 171
 development 171
 initial consonant clusters 171
 Pali influences 171
 voiceless sonorants 171
 vowels 171
 as tone language 172
 use of 170
 verbal complex 173
 voiceless nasals 171–172
 vowels 172
 see also Lolo-Burmese languages
- Buru
 speaker numbers 690
 see also Malukan languages
 Burundi
 Kinyarwanda 604
 Swahili 1026
 Burushaski 175–180
 case forms 176–177
 classification 249
 consonants 176
 diminutives 176
 double argument indexing 178
 Hunza dialect 175
 speaker numbers 175
 influence from other languages 179
 intransitive verbs 178
 Nagar dialect 175
 speaker numbers 175
 noun classes 176
 numerals 177
 plurals 176
 retroflexion 176
 subordinate clauses 178–179
 use of 175
 bilingualism 175–176
 speaker numbers 175
 verbs 177–178
 vowels 176
 word order 178
 SOV 178
 Yasin dialect 175
 speaker numbers 175
 syntax 178
 see also Austroasiatic languages; Dardic languages; South Asian languages
- Buryat 722
 use of 723
 see also Mongol languages
 Buschmann, Johann Carl 1140
 Butam 841
 see also East New Britain languages
 Buxinhua 729
 see also Mon-Khmer languages
 Byzantine Greek, Romani influences 898–899
- C**
- Cabecar 653
 Cacaopera
 classification 711
 see also Misumalpan languages
 Caddo 749
 see also Caddoan languages
 Caddoan languages 749
 classification 252
 grammars (books) 181
 history 181
 nouns 181
 phonemes 181
 scholarship 181
 sentence structure 181
 structure 181
 verbs 181
 see also Arikara; Wichita
 Cadorine
 classification 894
 see also Ladin
 Cahuapanan languages 41
 see also Andean languages
 Cambodia
 Bahnaric languages 725–726
 Katuic languages 726–727
 Khmer (Cambodian) 597
 Mon-Khmer languages 725
 Pali 830
 Pearic languages 728
 Cambodian *see* Khmer (Cambodian)
Cambridge History of the English Language 355
 Camden, William, Pictish 856
 Cameroons
 Adamawa-Ubangi languages 771
 Fulfulde 430
 Kanuri 578
 Mambila 691
 Camling, compound verbs 996–997
 Campa languages
 predicate structure 60
 use of 59
 see also Arawak languages
 Campbell, L., Nostratic 653–654
 Canaanite 932, 933
 see also Semitic languages
 Canada
 Cree 261
 Dutch 307
 Estonian 377
 Fijian 412
 Finnish 413
 French 427
 Italian 545
 Michif 709
 Nuuchahnulth 788
 Candoshi languages 41
 see also Andean languages
 Canonical forms, sign language morphology 942*f*
Cantiga da Ribeirinha 883
Cantiga de Garvaia 883
Cantiga d'Escárnio 883
 Capell, A 84
 Cape Verdean Creole 182–183
 classification 249*t*
 history 182
 influence from other languages 182
 lexicon 182
 reduplication 182
 use of 182
 see also Creoles; Pidgins
 Cape Verde Islands, Portuguese 883
 Carapana
 accent/tone 1096
 case markers 1096
 consonants 1094*t*
 speaker numbers 1092*t*
 verbs 1099–1100
 word order 1096
 see also Tucanoan languages
 Caretaker language, North American native languages 757
 Cariban languages 750
 adverbs 186
 class-changing 186
 classification 183, 185–186
 comparative studies 183
 gender 184–185
 lexicon 187
 morphemes 184–185
 morphology 184
 negation 186
 person-marking prefixes 185–186
 phonology 183
 possessives 184–185, 186
 postpositions 186
 reduplication 184
 semantics 187
 stops 183–184
 subordinate clauses 186–187
 suffixes 184
 syntax 186
 use of 40, 184*f*
 vowels 183–184
 weight-sensitive stress 183–184
 word order 186
 OVS 186
 workers in 183
 see also Akuriyo; Andean languages; Arara
 Carib languages
 classification 252–253
 Mapudungan language affiliations 701

- Carochi, Horacio 745
- Case
in agglutinating languages 416
European linguistic area 398, 398*f*
- Case markers
Abkhaz 1–2
Bikol 159*t*
Burmese 173
Carapana 1096
Desano 1096, 1097
Guaraní 468
Macuna 1097
Pisamira 1097
Retuara/Tanimuca 1096
Samar-Leyte 916, 916*t*
Siona 1097
Siriano 1096
Sumerian 1023, 1024*t*
Tatuyo 1096, 1097
Tucanoan languages 1096
Tuyuca 1097
- Casiguran Dumagat Agta
phonology 784
see also North Philippine languages
- Castrén, Matthias Alexander
Altaic languages 30
Mongol languages 723
- Catalan 188–192
demography 188, 190*t*
dialects 189*f*, 190
genetic relationship 188
geography 188
history 190
Latin 190
literature 190–191
post-World War II 191
Occitan *vs.* 799
as official language 191
phonology 188–190
sociolinguistics 191
typological features 188
use of 188, 190*t*
Andorra 188
France 188
Italy 188, 545
number of speakers 188
Spain 188
vocabulary 190
see also Indo-European languages; Portuguese;
Romance languages; Spanish
- Catawban languages, Siouan languages *vs.* 972
- Categoricals, nonobligatory, Southeast Asian
languages 1011*t*
- Categories, neutrality, inflection *see* Inflection
- Caucasian languages 192–197
classification 251
consonants 193, 193*t*
influence on other languages, Ossetic 814
kinships 196
morphology 194
nominative/absolutives 195
phonemes 193
phonetics 193
phonology 193
stress 194
postposition 195
Svan dialects 193, 193*t*
syntax 195
use of 192
verbs 195
vowels 193, 193*t*, 194*t*
word order 195
see also Abkhaz; Andi; Archi; Avar; Georgian;
Lak
- Caucasian *Sprachbund* 392–393
- Caudmont, Jean 226
- Cauqui *see* Aymaran
- Causatives *see* Inflection
- Cayapa *see* Barbacoan languages
- Cayuga languages
laryngeal features 543
stress 543
use of 543
see also Iroquoian languages
- Cayuse 750
see also Penutian languages
- Cebuano 197–199
affixes 198, 199
consonants 197–198
deictics 198
demonstrative pronouns 198
dictionaries 197
glottal stops 197–198
grammars (books) 197
history 197
phonology 197–198
Tagalog *vs.* 197
use of 197
verbs 198
vowels 197–198
see also Austronesian languages; Samar-Leyte;
Tagalog
- Cedilla, French orthography 428
- Celtiberian
classification 199–200
see also Celtic, Continental
- Celtic 199–201
classification 199–200, 251–252
genetic classification 246
Continental 199–200
history 199
influence on other languages, Old
English 358
Insular 200
reconstruction 246
Tocharian 1070
Welsh vocabulary 1170
see also Breton; Cornish; Goidelic languages;
Scots Gaelic; Welsh
- Central African Republic
languages
Adamawa-Ubangi languages 771
Fulfulde 430
official languages
French 917
Sango 917
- Central and Eastern Oceanic
languages 99–100
- Central German 445
- Central Luwian 37–38
- Central Malayo-Polynesian (CMP) 685
see also Malayo-Polynesian languages
- Central Semitic languages 931
imperfective 931
see also Semitic languages
- Central Siberian Yupik 201–204
enclitics 203
Greenlandic *vs.* 202
nouns 203
postbases 201–202
concatenative 203
in derivational morphology 202, 203*t*
lexical category changing 203
productivity 202–203
recursion 203
syntax interaction 202, 203
variable order 202, 203
verb derivation 202
verb 203
see also Algonquin languages; Arabic;
Arabic, as introflexing language;
Caddoan languages; Crow;
Eskimo-Aleut languages; Lakota;
Morphological Types; Nahuatl;
Ngan'gi; Ritwan languages; Tiwi
- Central Solomons languages 841
classification 204
gender 205
numbers 205
pronouns 205
reduplication 205
serial verb constructions 205
word order 205
see also Papuan languages
- Central Sudanic
classification 773
use of 775*f*
see also Nilo-Saharan languages
- Central West Greenlandic 1172
see also West Greenlandic
- Control Colombiano de Estudios de Lenguas
Aborígenes (CCELA) 230–231
- Ceremonial speech, Pitjantjatjara 871
- Chacha 41
- Chachi *see* Barbacoan languages
- Chad
Adamawa-Ubangi languages 771
Arabic 42
Kanuri 578
Nilo-Saharan 774
- Chadic languages 12, 206–208
classification 250
dictionaries 206
downstep 206
grammars (books) 206
ideophones 207
morphology 206
negation 207
noun-phrase syntax 207
noun pluralization 206
phonology 206
pluractional verbs 207
reduplication 207
syntax 206
as tonal language 206
use of 206
verbs 206
vowel systems 206
VSO 207
see also Africa; Africa, as linguistic area;
Afroasiatic languages; Cushitic languages;
Hausa
- Chaghatay 1053
- Chalas-KuRangal 282
see also Pashai languages
- Chalchiteko 705–706
official recognition 705–706
speaker numbers 706*t*
- Chaldean (Nestorian) Church 1033
- Chamberlain, Alexander
Chico language studies 225
Ryukyuan 908
- Chamicuro
pronominal suffix loss 60
see also Arawak languages
- Chang 968–969
see also Konyak languages
- Channel Islands, French 427
- Cha'palaachi *see* Barbacoan
languages
- Character signs, sign languages grammatical
comparisons 957–958
- Chatino 751
classification 819–821
speaker numbers 1213
syllable onsets 821–822
see also Oto-Mangean languages
- Chayama *see* Cariban languages
- Chayma 185*f*
- Chechen 194, 194*t*
see also Caucasian languages
- Chedepo 624
see also Grebo languages
- Chemakum 210
morphology 210
phonology 210
typology 210
- Cheremis *see* Mari languages
- Cherokee
classification 252
noun incorporation 544
tone 543
use of 542
see also Iroquoian languages
- Cheyenne
classification 25
stress 26
see also Algonquin languages
- Chhong 728
see also Pearic languages
- Chiapanec-Mangue languages 751
see also Oto-Mangean languages

- Chibchan languages 750
 classification 224, 252
 external relationships 209
 origins/development 209
 speaker numbers 208–209
 subgrouping 209
 types 208–209
 use of 40
see also Andean languages
- Chibchan-Paezan hypothesis 651–652
- Chi-Chewa *see* Nyanja
- Chichimeca Jonaz 751
see also Otopamean languages
- Chichimeko, classification 819–821
- Chickasaw 738–739
 consonants 739, 739*t*
 influence on other languages, Mobilian Jargon 716–717
 verbs 741
- Chickasaw-Choctaw trade language *see* Mobilian Jargon
- Chico languages 224–238
 classification 224, 230*t*, 231*t*, 233*t*
 demography 224, 227, 227*f*, 228*f*
 regional classification 229, 229*f*
 dialects 224–225
 historical studies 224, 225
 Caudmont, Jean 226
 Chamberlain, Alexander 225
 comparative studies 225
 cultural effects 225
 Loewen, Jacob 226
 Pinto, Constancio 226
 SIL 226
 history/development 224
 present study 229
 Centrol Colombiano de Estudios de
 Lenguas Aborígenes (CEELA)
 230–231
 use of 224
 geographical distribution 224
 Panama 224
 speaker numbers 224
 vocabularies 225
 workers in
 Caudmont, Jean 226
 Chamberlain, Alexander 225
 Loewen, Jacob 226
 Pinto, Constancio 226
- Chikomulselteko
 geographical distribution 705
see also Mayan languages
- Children, sign language acquisition 945
- Chile
 Andean languages 41
 Aymara 108
 Mapuche 41
 Mapudungan languages 701
- Chimakuan languages 210–211
 diminutives 210
 morphology 210–211
 phonology 210
 typology 210
 Wakashan languages 750
see also Chemakum
- Chimariko languages 750–751
 classification 505
see also Hokan languages
- Chimbu *see* Kuman (Chimbu)
- Chimbu-Wahgi languages
 classification 1086–1087
 geographical distribution 669, 670*f*
 pronouns 1087–1089
 verb root 1087
see also Trans New Guinea languages
- Chimchimeko *see* Oto-Mangean languages
- Chimila *see* Chibchan languages
- Chin (Tiddim: Tedim) 968–969
see also Kuki-Chin languages
- China
 Burmese 170
 Evenki 405
 Kazakh 588
- Khmuic languages 727
- Kirghiz 610
- Mang languages 729
- Mon-Khmer languages 725
- Palaung-Wa languages 727
- Palyu 729
- Sino-Tibetan 968
- Tai-Kadai 105
- Tai languages 1039
- Tibetan 1060–1061
- Tungusic languages 1103
- Uzbek 1145
- Wa 1155
- Waic languages 728
- Chinantec 211–213
 ‘ballistic stress,’ 212
 classifier 212
 ‘controlled stress,’ 212
 inflection 211–212
 nouns 212
 roots 211–212
 sandhi 212*t*, 213
 stem inflexion 212*t*
 stem modification 211–212
 stress 212
 tonal features 212
 tone sandhi 213*t*
 verbs 211–212, 212*t*
 prefixes 212
 words 211–212
see also Oto-Mangean languages
- Chinantecan languages 751
 progressive 212*t*
 VSO 211
see also Oto-Mangean languages
- Chinanteko 819
 time depth 819
see also Oto-Mangean languages
- Chinese 213–221
 classification 214*f*
 non-Mandarin group 214*f*
 dimorphic words 221
 distribution 214
 grammar 216
 comment 216
 topic 216
 identifiable morphemes 222
 overlapping exponence 223
 phonological form invariance 223
 suffixes 222, 222*t*
 influence on other languages
 Japanese 557
 Korean 615–616
 Lao 640
 Tocharian 1070
 Vietnamese 248, 728–729, 1149
 Wa 1156
 information processing 219
 as isolating language 221–224
 classifiers 221
 Mandarin *see* Mandarin Chinese
 marking 222
 Min group 219
see also Fuzhou
- monomorphemic words 221
 affixes 221
 classifiers 221
 human pronouns 221
- morphemes/word 222
- phonology 215
 finals 215*f*
 initials 215
 intonations of utterances 215–216
 syllables 215*f*
 tones 215
 pragmatics 216
 self-denigration 216–217
 script *see* Chinese script
 verbs, inflectional suffixes 221
- Wu group 219
see also Shanghai Chinese
- Xinjiang 1142
- Yue group 218
see also Hong Kong Cantonese
- see also* Arabic, as introflecting language;
 Central Siberian Yupik; Classification (of
 languages); Finnish (Suomi); Morphological
 Types; Polysynthetic languages
- Chinese script 217
 development 217
 from pictographs 217
hanzi 217
 reform of 217
 semantic character formation 217–218
 strokes 217
- Chinese Sign Language, finger/thumb negation
 957–958
- Chinook 750
see also Penutian languages
- ChiNyanja *see* Nyanja
- Chipaya 752
see also Native American languages
- Chiquitano
 classification 666, 666*t*
 morphology, word order 667–668
 use of 666–667
see also Macro-Jê languages
- Chiragh Dargwa 194, 194*t*
see also Caucasian languages
- Chitimacha 749
see also Muskogean languages
- Chitral languages 282
see also Dardic languages
- Chiwere 749
see also Siouan languages
- Chocho 751
see also Popolocan languages
- Chochoan languages 819–821
see also Oto-Mangean languages
- Choco languages
 classification 252–253
 use of 40
see also Andean languages
- Choctaw 738–739
 classification 252
 influence on other languages, Mobilian Jargon
 716–717
 noun phrases 740–741
 phonology, consonants 739, 739*t*
 verbs 740, 741
 word order 741–742
see also Muskogean languages
- Choctaw-Chikasaw 749
see also Muskogean languages
- Ch’olan 705–706
 speaker numbers 707*t*
see also Mayan languages
- Chon 752
see also Native American languages
- Chono 41
see also Andean languages
- Chontal languages 504
 classification 506
 speaker numbers 707*t*
see also Hokan languages
- Chorasmian 238–239, 538
 classification 251–252
 definite articles 540
 dual 540
 genders 238–239, 540
 imperfect tense 541
 modal forms 542
 palatal affricates 540
 past tenses 541
 phonology 238
 possessives 238
 pronouns 239
 script 238
 use of 238
 verbs 239
 vowels 238
see also Iranian languages
- Chorotegan
 time depth 819
 tone 821
- Ch’orti 705–706
 speaker numbers 706*t*
see also Mayan languages

- Chrau 726
 see also Bahnaric languages
 Christaller, Johann Gottlieb 17–18
 Christianity, Hebrew, study of 484
 Chugani 282
 see also Pashai languages
 Chuj 705–706
 dialects 705–706
 positionals 707
 speaker numbers 706*t*
 see also Mayan languages
 Chujean 705–706
 see also Mayan languages
 Chukchi (Chukot)
 classification 239
 male *vs.* female phonology 239–240
 speaker numbers 239–240
 see also Chukotko-Kamchatkan languages
 Chukot *see* Chukchi (Chukot)
 Chukotko-Kamchatkan languages 239–241
 circumfixes 240
 classification 251
 as endangered languages 239–240
 special characteristics 240
 use of 239
 speaker numbers 239–240
 vowel harmony 240
 see also Alutor; Language endangerment
 Chumashan 750–751
 see also Hokan languages
 Church Slavonic 241–243
 definition 241
 local varieties 241–242
 origins 241
 revisionism 242
 Russian, influence on 905
 see also Balto-Slavic languages; Bulgarian;
 Macedonian; Old Church Slavonic
 Chut languages 728–729
 see also Arem; Viet-Muong languages
 Chuvash 243–246, 1109
 consonants 244
 development 1109
 dialects 245
 distinctive features 244
 grammar 244
 lexicon 245
 nominative case 245
 origin/history 243
 phonology 244
 possessives 244–245
 related languages 243
 sound harmony 244
 use of 243
 verbs 245
 vowel harmony 245
 linguistic assimilation 244
 vowels 244
 written language 244
 see also Altaic languages; Turkic languages
 Cilappatrikaram 1047–1048
 Circum-Baltic linguistic area 64, 392–393
 Circumfixes, diachronic origins 287
 Circumflex, French orthography 428
 Circumstantial case *see* Case
 Circumstantial, linguistic areas 62
 Cladistics, Indo-European languages 529
 Classification (of languages) 246–257
 Central America 252
 see also Chibchan languages; Totonacan languages
 diffusion 246
 innovation spread 246–248
 plural markers 248
 sprachbund 248
 tense markers 248
 vocabulary borrowing 248
 word order 248
 genetic classification 246
 ancestral languages 246
 inflections 246
 subgroups 246
 tree diagrams 246
 geographical distribution 246, 247*f*
 grouping status 250
 index of synthesis 731
 isolates 249
 lexicostatistics 248
 African languages 248–249
 American languages 248–249
 ‘basic words’ 248
 morphological technique 731
 see also Morphological types
 North America 252
 see also Algonquin languages; Caddoan languages; Hokan languages; Iroquoian languages; Keres; Muskogean languages; Na-Dene languages; Ritwan languages; Salishan languages; Siouan languages; Wakashan languages
 Nostratic theory 249
 pidgins/Creoles 249
 relational concepts 731
 South America 252
 see also Arawak languages; Panoan languages; Quechua languages; Tucanoan languages; Tupian languages
 see also Mongolia; Morphological types
 Classifiers
 Algonquin languages 28
 Arawak languages 60
 Baniwa 61
 Chinantec 212
 Chinese 221
 Cubeo 1098
 Desano 1098
 Gondi 457
 Hmong 1013
 Hokan languages 508
 in isolating languages 221
 Karen 581
 Khmer (Cambodian) 599
 K’iche’an 708
 Korean 615
 Koreguaje 1098
 Kwakwala 1160*t*
 Lao 639–640
 Mandarin Chinese 1013
 Mayan languages 708
 Na-Dene languages 743–744
 Nepali 764
 Ngan’gi 766
 Oto-Mangean languages 823
 Palikur 61
 Persian, Modern 850
 Popti’ 708, 708*t*, 709*t*
 Q’anjob’al 708, 708*t*
 Retuara/Tanimuca 1098
 Ritwan languages 28
 Secoya 1098
 sign language 943, 943*f*, 952
 Siriano 1097, 1098*t*
 South Asia 997–998
 Southeast Asian languages 1013, 1014*t*
 South Philippine languages 1004
 Tajik Persian 1042
 Tamambo 1047
 Tariana 61, 1051
 Thai 1060
 Tucano 1098
 Tucanoan languages 1097, 1098, 1098*t*, 1099
 Tupian languages 1108
 Tuyuca 1098
 Vurës 1154
 Wakashan languages 1159–1160, 1160*t*
 Wolof 1185–1186
 Yucatecan 708
 Clicks
 Bantu languages 1017–1018
 Khoekhoe 602*t*
 Khoesaa languages 602, 602*t*
 Pidgins 859
 Southern Bantu languages 1018
 Xhosa 1018
 Zulu 1215–1216
 Clitic(s)
 ordering 130
 resumptive compounds 124
 see also Affixation
 Cluster maps, European linguistic area 402–403, 403*f*
 Coahuilteco 751
 see also Native American languages
 Coast Salish 749
 Coatlán 714
 see also Mixe-Zoquean languages
 Codex Leningradensis 483
 Codrington, R. H., Austronesian languages 98
 Coeur d’Alene 749
 see also Salishan languages
 Cofán 41
 see also Andean languages
 Cognitive semantics *see* Classifiers
 Cohen, M. 13
 Colima *see* Cariban languages
 Colonialism, Later Modern English development 344, 349–350
 Colorado *see* Barbacoan languages
 Columbia
 Andean languages 40
 Arawak 59
 Cariban languages 40
 Chibcha 40
 Choco languages 40
 Emberá 224
 Guajiro 59
 Palenquero 828
 Quechua 891
 Tucanoan languages 1091
 Waunmêu 224
 Comecrudan 751
 see also Native American languages
 Comelico 894
 see also Ladin
 Come to have verb, Southeast Asian languages 1015
 Comitative-instrumental syncretism, European linguistic area 399, 400*f*
 Common standard language, German *see* German
 Communication, Later Modern English development 344
 Comparatives, in introflexing language 52
 Complex sentence(s) 989
 Complex sentences
 Dravidian languages 300
 Pitjantjatjara 873
 Somali 989
 Compounding, sign language morphology 949, 949*f*
 Computer-supported writing *see* Writing/written language
 Con 728
 see also Lametic languages
 Concatenative postbases, in polysynthetic languages 203
 ‘Concentric circles’ model *see* World Englishes, ‘concentric circles’ model
 Conceptual blending *see* Lexical semantics
 Concord
 African-American Vernacular English (AAVE) 336
 Afrikaans 9
 American English 336
 Balkan linguistic area 125
 Bantu languages 140
 Benue-Congo languages 151
 Cushitic languages 274–275
 Domari 296
 English 343–344
 Fulfulde 432
 Gikuyu (Kikuyu) 450, 451*t*
 Gur languages 473
 Hurrian 515–516
 Kru languages 624
 Kwa languages 632
 Lithuanian 647–648
 Luganda 658
 Romanian 900
 Scots 925

- Concord (*continued*)
 Southern Bantu languages 1018
 Spanish 1021
 Swahili 1027–1028
 Telugu 1055–1056
 Tibetan 1061
 Toda 1072
 Togo Mountain languages 632
 Torricelli languages 1078
 Trans New Guinea languages 842
 Wambaya 1162
- Congo, Democratic Republic of
 Fulfulde 430
 Kinyarwanda 604
 Swahili 1026
- Conjugated prepositions, Old Irish 453
- Connacht, Irish, development of 454
- Conoy (Piscataway) 24
see also Algonquin languages
- Consonant(s)
 in agglutinating languages
 deletion 419
 gradation 418
 roots, in introflexing language 50
see also specific languages
- Continental Celtic *see* Celtic, Continental
- Converbs
 Amharic 35
 Assamese 995
 Cushitic languages 275
 Ethiopian linguistic area (ELA) 380
 Ethiopian Semitic languages 383
 Evenki 406–407
 Hindi 995
 Kannada 995
 Kazakh 590
 Nilo-Saharan languages 774
 Oriya 995
 Santali 995
 South Asian languages 995
 Tamil 995
 Tatar 1054
 Tigrinya 1064–1065
 Türkmen 1119
 Uzbek 1147
 Yukaghir 1211–1211
- Convergence, Indo-European language
 classification 530
- Coos 750
see also Penutian languages
- Copainalá
 aspects 714
 cliticization 714
 nouns 714
 phonology 713
 syllable cods 713
 word order 715
see also Mixe-Zoquean languages
- Copi 1018
see also Inhabane languages
- Coptic 38–40
 classification 250
 SVO 39
see also Afroasiatic languages
- Corachol 1140
see also Uto-Aztecan languages
- Cornish 257–258
 Breton *vs.* 257
 classification 200, 251–252
 revival/survival 258
 Welsh *vs.* 257
 workers in 258
see also Breton; Brythonic Celtic; Celtic; Pictish;
 Welsh
- Coroado *see* Puri (Coroado)
- Correlative
 Bengali 148
 Dardic languages 284
 Kashmiri 583–584
 Marathi 704
 Ossetic 818
 Pali 831
 Persian, Old 853
 Corsica, Italian 545
- Cotoname 751
see also Native American languages
- Cowlitz 749
see also Salishan languages
- Creativity, multilingualism 368
- Cree 258–263
 classification 24–25
 derivation 259
 primary stems 259
 recursive suffixation 259–260
 secondary 259
 dialects 261
 dictionaries 261
 grammars (books) 261
 inflexion 258
 language shift 319
 Michif, influences on 710
 noun incorporation 260
 incorporative verbs 260
 medials 260
 paradigmatic sets 260
 number 258
 associative plural constructions 259
 origin/development 28
 phonology, stress 26
 speaker numbers 26
 use of
 bilingualism 261
 Canada 261
 verbs 260
 inflexion 258
 parallel constructions 260
 word order 260
see also Algonquin languages; Ritwan
 languages
- Creek 749
 augmentative 264
 auxiliary verbs 266
 classification 252
 consonants 263
 dialects 754
 diminutives 264
 future 265
 glides 263
 history 263
 morphology 264
 agreement type 740–741
 nouns *see below*
 verbs *see below*
 noun morphology 264
 case marking 264
 creation from verbs 264
 number 264
 possession 264
 noun phrases 266
 phonology 263
 pitch-accent 739–740
 possessives 264
 postpositions 741
 resources 267
 resultative 264–265
 SOV 266
 suffixes 740
 syntax 266
 verb morphology 264
 infection 264–265
 negative statements 266
 plurals 266
 vowel length 740
 vowels 263–264
see also Mobilian Jargon (Mobilian);
 Muskogean languages
- Creole Portuguese
 Afrikaans, influence on 9, 11
 future 868
- Creoles 857–864
 classification 858
 lexical affiliation 858
 as continuum 866
 decreolization 869
 definition 857
 ergativity 862
 future languages 863
 gender markings 862
- influences from other languages,
 Portuguese 883
- lectallectal variation 867
- lexical semantics 866
- myths about 864
- noun classes 862
- origins/development 859
 Bickerton, D 861
 bioprogram hypothesis 861
 children *vs.* adults 859
 diffusion theory 860
 relexification 860
 sociohistorical context 863
 source morphemes 862
 substrate theory 859
 superstrate theory 860
 universals theory 860
- Pidgins *vs.* 862
- shared features 861–862
- sign languages 956
- tense-mood-aspect systems 862
- types 862
 USA 1127
- variations in 865
 acrolect 868
 basilect 867
 mesolect 868
- see also* African-American Vernacular
 English (AAVE); Austronesian
 languages; Bislama; Cape Verdean Creole;
 English, nonnative; Fanagalo; Gullah;
 Hawaiian Creole English (HCE); Krio;
 Louisiana Creole; Mobilian Jargon
 (Mobilian); Morrobalama; Palenquero;
 Pidgins; Russenorsk; Sango; Tiwi; Tok
 Pisin; Yanito
- Creole theory, African-American Vernacular
 English development 337
- Crimean Gothic 460
- Crimean Tartar 1109
see also Turkic languages
- Critical Period Hypothesis (of language
 acquisition), sign languages 945
- Croatia
 Hungarian 514
 Romanian 901
- Croatian 936
 Slovene *vs.* 981
- Croatian-Bosnian-Serbian Linguistic complex
see Serbian-Croatian-Bosnian Linguistic
 Complex
- Croatian-Serbian-Bosnian Linguistic complex
see Serbian-Croatian-Bosnian Linguistic
 Complex
- Cross River languages 151
see also Benue-Congo languages
- Crow 749
 active verbs 269
 augmentative 269
 classification 252
 consonants 267, 267t
 diminutives 269
 final markers 269
 habitual 269
 morphology 268
 morphosyntax 269
 noun phrases 269
 object incorporation 269
 orthography 267
 Crow Agency Bilingual Education
 Program 267
 phonology 267
 plurals 268
 possessors 268
 postpositions 269
 speaker numbers/location 971–972
 stops 267–268
 subordinate clauses 269
 suffixes 269
 switch-reference 269
 use of 267
 speaker numbers 267
 verbs 268–269
 vowels 268, 268t

- see also* Central Siberian Yupik; Lakota;
 Language endangerment; Omaha-Ponca;
 Siouan languages
 Crow Agency Bilingual Education
 Program 267
 Crowther, Samuel Ajayi, Yoruba 1207
 Cua
 speaker numbers 726
see also Bahnaric languages
 Cubeo
 adjectives 1099
 classification 1091
 consonants 1094
 evidentiality 1100
 morphemes 1096
 nasalization 1095
 noun classifiers 1098
 speaker numbers 1092*t*
 verbs, evidentiality 1100
 word order 1096
see also Tucanoan languages
 Cuicatec *see* Mixtecan languages
 Cuitlatec 748
 Spanish, influences of 651–652
see also Native American languages
 Culavamsa, Pāli commentaries 832
 Culli 41
 Culture
 dislocation 325
 European linguistic area 390
 Nostratic theory 786
 Cuna *see* Chibchan languages
 Cuneiform script
 Elamite 316
 Hittite 502
 Old Persian 852
 Cuoi languages 728–729
see also Viet-Muong languages
 Cupeño 270–272
 as agglutinating language 270
 auxiliary complexes 271
 classification 252
 dual agreement marking 271
 ergative 271
 imperfective 271
 nouns 270–271
 animate 270–271
 inanimate 270–271
 split-ergative system 271
 syntax 271
 use of 270
 USA 270
 word classes 271
 word order 271
see also Takic languages
 Cushitic languages 12, 272–276
 Amharic, influences on 33
 case system 274–275
 classification 250
 concord 274–275
 converbs 275
 genders 274–275
 Highland East Cushitic *see* Highland East
 Cushitic (HEC) languages
 imperfective 274–275
 iterative 274–275
 morphology 274
 noun plurals 274–275
 number of speakers 272–273
 phonology 274
 postpositions 274–275
 relationships 273*f*
 SOV 275
 syntax 275
 tense-mood-aspect 274–275, 275*f*
 types 273
 use of 272–273
 verb person marking 274–275, 274*f*, 275*f*
see also Afar (Qafar); Afroasiatic
 languages; Agaw; Chadic languages;
 Ethiopian linguistic area (ELA);
 Highland East Cushitic (HEC)
 languages; Omotic languages;
 Oromo; Somali
 Cyprus
 Modern Greek 464
 Turkish 1112
 Cyrillic alphabet
 Abkhaz 1
 Azerbaijanian 111
 Belorussian 147
 Kazakh 589
 Kirghiz 611
 Macedonian 664
 Tajik Persian 1041–1042
 Turkmen 1117
 Ukrainian 1122
 Uyghur 1143
 Uzbek 1146
 Yakut 1200
 Czech 276–277
 adjectives 277
 cardinal numbers 277
 classification 251–252, 974–975
 dialects 276
 future 277
 grammatical gender 277
 imperfective 277
 inflectional morphology 277
 nouns 277
 as official language 276
 origins 276
 phonemes 276–277
 pro-drop 277–277
 pronouns 277
 verbal morphology 277
see also Balto-Slavic languages; Slavic
 languages; Sorbian
 Czech Republic
 Czech 276
 German 444
 Slovak 977
D
 Daco-Romanian, Romanian dialect 901
 Dagaari 472
see also Oti-Volta languages
 Dagan languages 1087
see also Trans New Guinea languages
 Dagbani 472
see also Oti-Volta languages
 Daghestanian languages 193
see also Caucasian languages
 Dagur 722
 use of 723
see also Mongol languages
 Dahlik
 use of 382–383
see also Ethiopian Semitic languages
 Dahl's Law
 Bantu languages 139
 Kinyarwanda phonology 607
 Daju
 classification 773
 speaker numbers 772–773
see also Nilo-Saharan languages
 Dakota 749
see also Siouan languages
 Dakota, language shift 319
 Dakotan *see* Lakota
 Damana *see* Chibchan languages
 Dameli
 case-marking 284
 classification 282
 sibilants 283
 speaker numbers 282
see also Kunar languages
 Damentch 282
see also Pashai languages
 Danao languages 1001–1002, 1002*t*
see also South Philippine languages
 Danau 727
see also Palaung-Wa languages
 Dani languages 1087
see also Trans New Guinea
 languages
 Danish 279–282
 adjectives 280
 classification 251–252
 consonants 279
 history 279
 Ancient Scandinavian 279
 Old Scandinavian 279
 influence on other languages
 Inuit 374
 West Greenlandic 1175
 language authorities 281
 morphology 280
 nouns 280
 as official language 279
 orthography 279
 personal pronouns 280
 possessives 280
 pronunciation 279
 questions 281
 sentence order 280–281, 281*t*
 stød 280
 syntax 280
 tones 280
 use of 279
 verbs 280
 vowels 279–280
see also Germanic languages; Icelandic;
 Norse, Old; Norwegian; Scandinavian
 languages
 Danube *Sprachbund* 392–393
 Dardic languages 282–285, 582
 agreement patterns 284
 aspiration loss 283
 case-marking 284
 classification 251–252, 282
 uncertainties 283
 correlative 284
 counting systems 284
 gender 284
 history/development 283
 influence from other languages 282
 morphosyntax 284
 phonology 283
 retroflex affricates 283–284
 sibilants 283
 tonal system 526–527
 use of 282
 Pakistan 282, 283
 speaker numbers 282
 vowels 526–527
 word order 284
 written forms 282
see also Brahui; Burushaski; Indo-Aryan
 languages; Indo-Iranian languages;
 Kashmiri
 Darra-i-Nur (upper and Lower)
 classification 282
see also Pashai languages
 Dative
 South Asian languages 997
 Standard Average European (SAE) languages
 393–394
 Day 3
see also Adamawa-Ubangi languages
 Deaf communities, deaf cultures, formation 940
 Decreolization, Creoles 869
 Décsy, G, European linguistic area 390–391
 Defaka, Ijò *vs.* 517
 Definite articles, Standard Average European
 (SAE) languages 393–394
 Defoid 151
see also Benue-Congo languages
 De'kwana
 use of 185*f*
see also Cariban languages
 Delafosse, Maurice, Mande language
 classification 696
 Democratic Republic of the Congo *see* Congo,
 Democratic Republic
 Demonstratives
 Karitiána 1106
 Kashmiri 584
 Mawé 1106
 Mekéns 1106

- Demonstratives (*continued*)
 Mundurukú 1106
 Ossetic 815
 Tucanoan languages 1099, 1099*t*
 Tupian languages 1106
 Tuyuca 1099, 1099*t*
- Dempwolff, Otto
 Austronesian languages 98
 Proto Oceanic 687
- Dengebu 613
see also Kordofanian languages
- Denial *see* Negation
- Denmark
 Danish 279
 German 444
 Greenlandic (Kalaallisut) 1172
- Dental fricatives, nonnative English 360
- Derivational morphology,
 postbases 202, 203*t*
- Desano
 accent/tone 1096
 adjectives 1099
 consonants 1094*t*
 grammar, case markers 1096
 limiting adjectives 1098
 morphemes 1096
 nasalization 1095
 noun classifiers, inanimate 1098
 speaker numbers 1092*t*
 syllable pattern 1095
see also Tucanoan languages
- Devanagari script
 Bengali writing systems 148, 524
 Indo-Aryan language writing systems 524
 Kashmiri 583
 Punjabi writing systems 886
- Dhivehi 285–287
 classification 251–252
 consonants 285
 as first language 285
 Islam, influence of 285
 morphology 286
 nouns 286
 as official language, Maldives
 Islands 285
 origin/development 285
 orthography 285
 Roman alphabet 286
 ‘Thaana,’ 285
 phonology 285
 Sinhala *vs.* 285
 syntax 286
 use of 285
 geographical distribution 285
 verbs 286
 vocabulary 285
 vowels 285
 word order 286
see also Dardic languages; Indo-Aryan
 languages; Indo-European languages;
 Sinhala
- Diachrony
 affixes 287
 ergative 287
 morphological types 287–293
- Diaguita 41
- Dialect(s)
 Arabic 55–56
 Ashkun 787
 Assamese 78
 definition 385
 Ijò 518
 Kati 787
see also specific languages
- Dida languages 623–624
see also Kru languages
- Dien (Diyari) 87
see also Australian languages
- Diffusion area *see* Linguistic areas
- Diffusion theory, Creole origins 860
- Diglossia, definition 323–324
- Digoron, classification 813
- Dimasa 968–969
see also Bodo-Koch languages
- Dime 805
see also Omotic languages
- Diminutives
 Arabic 52
 Bantu languages 1018–1019
 Berber languages 154
 Bikol 160
 Burushaski 176
 Chimakuan languages 210
 Creek 264
 Crow 269
 Dutch 203–204
 French 428–429
 Fulfulde 431
 Gikuyu (Kikuyu) 452
 Guarani 467
 Hokan languages 507–508
 in introflexing languages 52
 Kwakwaka 1159*t*
 Mande languages 698
 Native American languages 757
 Nuuchahnulth (Nootka) 789
 Nyanja 796–796
 Romance languages 896–897
 Ryukyuan 907
 Salishan languages 913
 Shona languages 938
 Sioux 755
 Spanish 1021
 Swahili 1027
 Tupian languages 1106
 Wakashan languages 1159
 Waray-Waray 915–916
 Wolaitta 1180
 Xhosa 1188*t*
 Yanito 1202
 Yiddish 1204
- Dimorphic words, in isolating language 221
- Dinka 293–294
 classification 253
 SVO 294
 tones 294
 use of
 speaker numbers 772–773
 Sudan 293
 vowels 293, 293*t*
 word order 294
 workers in, Andersen, Torben 293
see also Nilo-Saharan languages; Nilotic
- Dipavamsa, Pāli commentaries 832
- Directional verbs, Southeast Asian
 languages 1014
- Disappearing languages
 definition 319–320
 reversal/revitalization strategies 326
see also Language endangerment
- Discourse morphology, Native American
 languages 746
- Distributional criteria, word *see* Word
- Ditidaht 1157
 syntax, word order 1160
- Dixon, Robert M W
 Australian languages 250
 long-range comparisons 649
- Dixon, Ronald P, Hokan languages 504
- Diyari (Dieri) 87
see also Australian languages
- Dizoid 805
see also Omotic languages
- Djenne Chiini 990–991
see also Songai languages
- Djibouti 272–273
- Doabi Punjabi 886
see also Punjabi
- Dogon 771
 classification 253
 dialects 294
 imperfective 294
 noun class system 294
 tones 294
 use of
 Mali 771
 speaker numbers 294
 vowel harmony 294
 vowels 294
 word order 294
see also Adamawa-Ubangi languages; Niger-
 Congo languages
- Dogri
 classification 522
 tonal system 526–527
 vowels 526–527
see also Indo-Aryan languages
- Dolgopolsky, Aharon, Nostratic theory 249,
 653–654
- Dolores, Juan, Tohono O’odham research 1074
- Domari 295–297
 adjectives 296
 Arabic, influences from 295, 296
 classification 251–252
 concord 296
 consonants 295
 definition 295
 demonstratives 296
 fricatives 295
 history 295
 intransitive verbs 296
 morphology 296
 nominative *vs.* oblique 296
 nouns 296
 perfect 296
 person markers 296
 phonology 295
 possessives 296
 pronouns 296
 Romani *vs.* 295
 syntax 296
 tenses 296
 use of 295
 verbs 296
 vowels 295–296
see also Dardic languages; Indo-Aryan
 languages; Indo-European languages;
 Romani
- Downstep
 Bantu languages 139
 Chadic languages 206
 Efik 314
 Guugu Yimithirr 473
 Ijò 518
 Kanuri 578
 Kwa languages 632
 Luo 658–659
 Nilo-Saharan languages 774
 Omaha-Ponca 803
- Dravidian languages 297–307
 adjectives 298
 adverbs 298–299
 agreement 299, 299*t*, 300*t*
 classification 251
 Nostratic theory 249
 complex sentences 300
 consonants 298, 298*t*
 contacts
 Indo-Aryan languages 306
 Munda languages 737
 dative-subject sentences 300
 equational sentences 300
 expressives 299
 future 304
 habitual 304–305
 Nostratic theory 653–654, 786
 nouns 298, 301
 ablative 302
 accusative 301–302
 agglutinative morphology 301
 case suffixes 301
 genitive 302
 instrumental case 302
 nominative case 301
 numerals 303, 303*t*
 plural suffixes 301
 pronouns 302, 302*t*, 303*t*
 particles 299
 phonology 297
 postpositions 301, 302
 Sanskrit, influences from 920–921
 script *see* Dravidian script

- Sindhi 960–961
 syntax 298
 types 297, 297*t*
 Central 297
 Ellis, Francis Whyte 297
 North 297
 South 297
 use of, India 297
 verbs 304
 auxiliary verbs 306
 finite verbs 304
 imperative 305, 305*t*
 negative finite verbs 304
 nonfinite verbs 305
 personal suffixes 305, 305*t*
 tenses 304
 vowels 297, 298*t*
 word classes 298
 word order 299
see also Brahui; Elamite; Gondi; Indo-Aryan languages; Kannada; Kurukh; Malayalam; Romani; South Asian languages; Tamil; Telugu; Toda
- Dravidian script 297
- Dual, in introflecting language 52
- Duit *see* Chibchan languages
- Dunbavin, Paul, Pictish 855–856
- Durbin, M., Cariban language
 classification 183
- Duru languages 3
see also Adamawa-Ubangi languages
- Dutch 307–311
 Afrikaans *vs.* 8, 9
 alveolar consonants 308
 articles 308
 assimilation 308
 classification 251–252
 consonants 308
 declarative main clauses 309
 derivational prefixes 308–309
 derivational suffixes 308–309
 dictionaries 308
 diminutives 203–204
 elision 308
 genders 309
 genetic relationships 307
 grammars (books) 308
 history 307
 Bible translation 308
 Old Dutch (Old Low Franconian) 307
 texts 307–308
 influence on other languages 310
 Afrikaans 7–8, 310
 colonial effects 310
 English 310–310
 Frisian 310
 Malayalam 680–681
 Saramaccan 858
 Sinhala 964
 Sranan 858
 morphology 308
 nouns 308
 orthography 309
 phonetics 308
 phonology 308
 regional/social variations 310
 Belgium 310
 dialects 310
 Flemish change 310
 sample sentence 309
 special characteristics 307
 stressed prefixes 309
 syntax 309
 use of 307
 verbs 308, 309
 vocabulary 309
 English 309
 French 309
 Latin 309
see also Afrikaans; English, Early Modern; English, Later Modern; English, Modern; German; Germanic languages; Papiamentu; Scots
- Dyirbal
 avoidance languages 91
 morphology 88–89
 syntax 88–89
see also Australian languages
- Dyren, Isidore, fields of work, Austronesian languages 98
- Dzhidezmo *see* Judeo-Spanish
- Dzubukua 667
see also Kariri (Kariri-Xocó)
- E**
- East Bird's Head (EBH) languages
 classification 1176
 Tense-Mood-Aspect verbal complex 1176
see also West Papuan languages
- Eastern Gurage languages 382–383
see also Ethiopian Semitic languages
- Eastern Kru languages
 Ivory Coast 623–624
 number of speakers 623–624
see also Kru languages
- Eastern Sudanic 775*f*
- Eastern Yiddish 1206
see also Yiddish
- East Luwian 37–38
see also Luwian
- East New Britain languages 841
see also Papuan languages
- East Syrian Church 58
- Ebang 613
see also Kordofanian languages
- Eblaité 313–314, 930
 grammar features 313
 sources 313
see also Afroasiatic languages; Akkadian; Semitic languages; Sumerian; Ugaritic
- Ebrie 631
- Ecclesiastical History of the English Peoples* 356
- Ecuador
 Andean languages 41
 Awá 224
 Aymaran 41
 Emberá 224
 Quechua 41
 Tucanoan languages 1091
- 'Eddic' poetry, Old Icelandic 779–780
- Edoid 151
see also Benue-Congo languages
- Edomite, Phoenician *vs.* 854
- Education
 Arabic 55
 language vitality role 324
 pidgins 867
- Efik 314–316
 Bantu *vs.* 314–315
 classification 314
 downstep 314
 as endangered language 314
 grammars (books) 314
 orthography 314
 tone 315
 focus 315
 grammatical characteristics 315
 lexical characteristics 315
 use of 314
see also Benue-Congo languages
- Efutu 631
see also Guang languages
- Ega 631
see also Kwa languages
- Egypt
 Aramaic 56
 Berber 152–153
 Modern Greek 464
- Egyptian 12, 38–40
 classification 250
 earliest record 12
 grammar 39
 hieroglyphics *see* Egyptian hieroglyphs
- Late 39
- Middle 39
 possessives 39
 VSO 39
 writing system 38, 39
see also Afroasiatic languages
- Egyptian hieroglyphs 38, 39
- Eipo (Eipomek) 1085–1086
see also Mek languages
- Eivo 841
see also North Bougainville languages
- Elamite 316–317
 classification 249
 earliest evidence 316
 morphology 316–317
 periods 316
 relation to other languages 316
 script 316
 syntax 316–317
 use of 316
 word order 316–317
see also Dravidian languages; Iranian languages; Persian, Old; Sumerian
- Elatives, in introflecting language 52
- Ellis, Francis Whyte, Dravidian 297
- El Salvador *see* Mayan languages
- Elvish 76
- Emberá
 classification 224
 historical studies 226
 Uribe, José Vincente 226
 use of 224
see also Choco languages
- Emilian *see* Italian
- Enclitics, in polysynthetic languages 203
- Endangered languages *see* Language endangerment
- Ende 420
see also Flores languages
- Endzelin, Jan
 Balto-Slavic languages 135–136
 Latvian 644
- Enets (Yenisey-Samoyed) 1129–1130
see also Samoyed languages
- Enga
 classification 1086–1087
 speaker numbers 836
 Sranan, influences on 858
see also Engan languages; Papuan languages; Trans New Guinea languages
- Engan languages
 classification 1086–1087
 geographical distribution 669, 670*f*
see also Samberigi (Sau); Trans New Guinea languages
- English
 African American *see* African-American English (AAE)
 African-American vernacular *see* African-American Vernacular English (AAVE)
 American *see* American English
 classification 251–252
 concord 343–344
 as global language 326
 influences from other languages
 Dutch 310–310
 French 248
 Greek 248
 Latin 248
 Mayan languages 708–709
 Scots Gaelic 927
 Welsh 1170
 influences on other languages
 Afrikaans 8
 Algonquin languages 24
 Bengali 148
 Bislama 162
 Dutch 309
 Fanagalo 412
 French 425, 429
 see also Franglais
 Gikuyu 449–450
 Gullah 470–471
 Hawaiian Creole English 481
 Hindi 495–496

- English (*continued*)
- Hong Kong Cantonese 218
 - Indo-Aryan languages 522
 - Inuit 374
 - Inupiaq 536
 - Japanese 557
 - Jèrriais 563
 - Korean 615–616
 - Krio 617
 - Malayalam 680–681
 - Mayan languages 708–709
 - Michif 710
 - Modern Standard Arabic (MSA) 43
 - Niuean 776
 - Punjabi 889
 - Saramaccan 858
 - Sinhala 964
 - Telugu 1055, 1058
 - Tok Pisin 1077
 - Yanito 1202
- long-range comparisons
- chance similarities 652
 - with German 651
 - sound correspondences 650–651
- as official language
- Fiji 413
 - Israel 485
 - Malta 688
 - Philippines 1035
 - Vanuatu 161
- possessives 340
- progressive 345–346
- Scots *vs.* 925
- spread 363
- stratification 363
- SVO 341
- use of, USA 1123
- see also* African-American Vernacular English (AAVE); Gullah
- English, Early Modern 339–343
- diversity 341
 - social differences 341–342
- grammar 340
- function word development 341
 - gender changes 340
 - nominal inflexion 340
 - periphrastic verb phrases 340
 - pronouns 340
 - third-person neuter possessive 340
 - verbal inflection 340
 - word order 341
- history 339
- educational opportunities 339
 - printing press introduction 339
- lexis/semantics 341
- French loanwords 341
 - Latin loanwords 341
 - technical terminologies 341
- phonology 339
- consonants 340
 - diphthong changes 339
 - Great Vowel Shift 339
 - short vowels 340
- regulation of variants 345
- see also* Dutch; English, Later Modern; English, Modern; Germanic languages; Middle English; Old English
- English, Later Modern 343–351
- 'be + -ing' construction 345–346
 - definition 343
 - double negatives 346–347
 - external history 343
 - American English development 344
 - colonization 344
 - communication developments 344
 - social mobility 344
 - transport changes 344
- inflection 344–345
- innovations 345
- lexical innovation 349
- colonial influences 349–350
 - rates 349*f*
 - scientific classification 349–350
- morphology 344
- personal pronouns 344–345
 - phonology 347
 - consonant dropping 347–348
 - consonants 347
 - Great Vowel Shift 349
 - lengthened vowel 348
 - rhotics 347
 - vowels 348
 - preposition stranding 346–347
 - pronouncing dictionaries 344
 - relativization 345
 - split infinitives 346–347
 - workers in 343
 - see also* Dutch; English, Early Modern; English, Modern; Middle English; Old English
- English, Modern 327–334
- accusative case 332
 - adjectives 330–331
 - adverbs 330–331
 - American–British differences 332
 - case structure 330
 - changes 332
 - clauses 331
 - complementation patterns 332
 - conjunctions 330–331
 - coordination 332
 - count nouns 330–331
 - determiners 330–331
 - influence from other languages 329–330
 - information structure 332
 - inserts 330–331
 - lexis 329
 - affixes 330
 - American–British differences 330
 - inflection lack 330
 - vocabulary size 329
 - mass nouns 330–331
 - modal auxiliaries 330–331
 - morphology 330
 - nominative case 332
 - noun phrases 331
 - orthography 327
 - American–British differences 328
 - spelling reform 328
 - vowels 327–328
- phonology
- American–British differences 329
 - changes 329
 - consonants 328
 - intonation 329
 - prominence/stress 328
 - received pronunciation 329
 - rhythm 328–329
 - voiced/voiceless consonants 328
 - vowels 328
- prepositions 330–331
- primary auxiliaries 330–331
- pronouns 330–331
- sentence structure 332
- standardization/prescriptivism 333
- media use 333
 - regional pronunciation 333
 - usage questions 333
- subordination 332
- syntax 330
- verbs 331
- lexical verbs 330–331
- word classes 330
- word order 330
- see also* Dutch; English, Early Modern; English, Later Modern; Germanic languages; Middle English; World Englishes
- English, nonnative 359–363
- borrowing 360
 - grammar 361
 - aspectual system 361
 - topic prominence 361
 - grammatical growth 361
 - inflection 361
 - lexical semantics 360–361
 - lexicon 360
 - morphology 360
 - native *vs.* nonnative speakers 360
 - native *vs.* nonnative varieties 359
 - phonetics/phonology 360
 - aspirated voiceless stops 360
 - dental fricatives 360
 - Singapore English 360
 - vowel inventory 360
 - register 361
 - British English (GB) 361
 - Singapore English (SIN) 361
 - Speak Good English* movement 361–362
 - stigmatization 361
 - theoretical approaches 362
 - variation measurement 360
 - see also* Creoles; Pidgins
- Englishization 365
- English Jewish 565
- development 568
 - writing system 568
- Eotilé 631
- see also* Tano languages
- Equative constructions, Standard Average European (SAE) languages 393–394
- Equatorial Guinea, Spanish 1020
- Ergative
- Afroasiatic languages 14
 - Andean languages 40
 - Apinajé 668
 - Arrernte 74
 - Australian languages 81
 - Bactrian 115
 - Basque 145–146
 - Cupeño 271
 - Jê languages 668
 - Kakutungu (Kalkutung) 88
 - Kariri (Kariri-Xocó) 668
 - Kashmiri 584
 - Maxakali 668
 - Morrobalama 735
 - Niuean 776
 - North Philippine languages 784
 - Nuristani languages 788
 - Obo Manobo 1005–1006, 1006*t*
 - Pama-Nyungan languages 88
 - Panará 668
 - Pitjantjatjara 872
 - South Philippine languages 1002
 - Tohono O'odham 1075
 - Wambaya 1163
 - Warlpiri 88, 1166–1167
 - West Greenlandic 1173–1175
 - Xokleng 668
 - Yalarnnga 88
- Eritrea
- Agaw 272–273
 - Bilen 272–273
 - Cushitic languages 272–273
 - Dahlik 382
 - Italian 545
 - Nilo-Saharan languages 774
 - Tigre 382
 - Tigrinya 382, 1063
- Erman, A, Hamitic theory opponent 13
- Erzya 1129–1130
- see also* Mordvin languages
- Esalen languages
- classification 506
 - see also* Hokan languages
- Eskimo-Aleut languages 748
- and Amerind 655
 - case marking 371–372
 - characteristics 371
 - classification 251, 1172
 - definition 371
 - genetic relationships 371
 - historical aspects 747–748
 - history 371
 - relations between 372
 - research history 371
 - Rask, Rasmus 371
 - sentences 371–372
 - SOV 371–372
 - verbs 371–372
 - vowels 371–372
 - word building 371–372

- word order 371–372
see also Aleut; Central Siberian Yupik;
 Greenlandic (Kalaallit); Inupiaq;
 Native American languages; Polysynthetic
 languages; West Greenlandic
- Esmeraldeño *see* Barbacoan languages
- Esperanto 76–77, 375–377
 accusatives 376
 aims 375
 as auxiliary language 75–76
 consonants 376
 as L2 375
 morphology 376
 opposition to 375
 origin/development 375
 World Esperanto Conference 375
 Zamenhof, Ludovic Lazar 375
 orthography 376
 plurals 376
 pronunciation 376
 Slavic influence 376
 SVO 376
 syntax 376
 tenses 376
 vocabulary 376
 vowels 376
 word order 376
 word stems 376
 Zamenhof, Ludovic Lazar 76–77, 375
- Esselen 750–751
see also Hokan languages
- Estonia 377
- Estonian 377–378
 classification 1129–1130
 consonantism 1130
 consonants 377
 dialects 377
 extrasegmental quantitative prosody 377–378
 history 377
 negation 1131
 as official language 377
 palatalized coronal consonants 378
 phonology 377
 use of 377
 verbs 1131
 vowels 377
 word order 1132
see also Finnic languages
- Etchemin 24
see also Algonquin languages
- Eteograms, Pahlavi (Middle Persian) 827
- Ethiopia 930
 Agaw 272–273
 Amharic 33
 classification 250
 Cushitic languages 272–273
 Fulfulde 430
 Hadiyya 272–273
 Highland East Cushitic (HEC) languages
 272–273, 488–489
 national language *see* Oromo
 Nilo-Saharan languages 774
 Sidamo 272–273
 Tigrinya 382, 1063
 Wolaitta 1179
see also Africa; Afroasiatic languages;
 Ethiopian linguistic area (ELA); Ethiopian
 Semitic languages; Semitic languages
- Ethiopian linguistic area (ELA) 64, 378–382
 converbs 380
 definition 378
 grammar 380
 ideophones 379
 imperfective 380
 lexicon 381
 phonology 379
 pharyngeal fricatives 380
 possessives 380
 postpositions 380
 research history 378
 Ferguson, C A 379
 Leslau, W 378–379
 Moreno, W W 378–379
 Zaborski, A 379
- SOV 380
see also Africa; Africa, as linguistic area;
 Amharic; Areal linguistics; Balkan
 linguistic area; Cushitic languages;
 Ethiopian linguistic area (ELA); Ethiopian
 Semitic languages; Europe, as Linguistic
 Area; Ge'ez; Ge'ez; Highland east Cushitic
 languages; Highland East Cushitic (HEC)
 languages; Omotic languages; Oromo;
 Somali; Tigrinya; Wolaitta
- Ethiopian Semitic languages 382–384
 classification 382
 converbs 383
 demography 382
 geographical origin 383
 ideophones 383
 imperfective 383
 morphology 383
 phonology 383
 possessives 383
 syntax 383
 use of 382
see also Afroasiatic languages; Amharic;
 Argobba; Ethiopian linguistic area (ELA);
 Ge'ez; Semitic languages
- Ethnolinguistic vitality *see* Language
 endangerment
- Ethnologue 384–388
 contents 385
 history 385
 Gordon, Raymond G Jr. 385
 Grimes, Barbara G 385
 Pittman, Richard S 385
 language identification 385
 workers in
 Gordon, Raymond G Jr. 385
 Grimes, Barbara G 385
 Pittman, Richard S 385
see also Language endangerment; SIL
- Étiemble, René 425
- Etruscan 388
 archaeological/historical context 388
 historical aspects 388
 inscriptions 388
 names 388
- Eurasian *Sprachbund* 392–393
- Europe, as linguistic area 388–405
 alienability 394–395
 approaches 389
 cultural 390
 geographical 389–390
 political 390
 simple languages 390
 case distinctions 398, 398f
 center *vs.* periphery approach 393
 Haspelmath, M 393
 cluster maps 402–403, 403f
 comitative-instrumental syncretism 399, 400f
 as contact-superposition zone 403
 egalitarian methods 390
 Décsy, G 390–391
 Haarmann, H 391
 historical overview 388
 isoglosses 394, 395
 geographical distribution 395
 morphology 398
 nominal cases 398–399
 ordinal derivation 400, 401f
 perfect 393–394
 phonology 395
 rounded vowels 395, 396f
 vowel length 397, 397f
 quintessence 402
 reduplication 401, 402f
 segregating approach 392
 Lewy, E 392
- EUROTYP project 388–389, 392–393
- Evenki 405–408
 agreement 406, 406t
 case 406
 contacts, Yakut 1200
 converbs 406–407
 dialects 405–406
 future 406
- modality markers 407
 morphemic ordering 407
 morphology 406
 negation 407
 non-finite verb forms 407
 number 406
 possessives 406
 possessivity 406
 resultative 406
 sentence structure 406
 SOV 406
 tense-aspect system 406
 use of 405
 valency change 407
 voice system 407
 writing system 405–406
see also Altaic languages; Tungusic languages
- Evidentiality
 Aymara 109
 Balkan linguistic area 130–131
 Cubeo 1100
 Karo 1108
 Kazakh 590
 Koreguaje 1100
 Panoan languages 834
 Quechua languages 892–892
 Retuara/Tanimuca 1100
 Secoya 1100
 Siona 1100
 Siriano 1101, 1101t
 Tariana 1051
 Tucanoan languages 1100, 1101t
 Tupian languages 1108
 Tuyuca 1100, 1101
 Uyghur 1144
 Uzbek 1147
 Wanano 1100
- Ewe 408–409
 consonants 408
 dialects 408
 double object constructions 409–409
 history 408
 ideophones 408
 morphology 408
 as national language 408
 nouns 408
 phonology 408
 sociolinguistics 408
 syntax 409
 tenses, lack of 409
 tone language 408
 use of 408
 speaker numbers 408
 verbs 632
 word order 409
see also Gbe languages
- Experiencers, Standard Average European (SAE)
 languages 393–394
- Extinct languages
 definition 319–320
 reversal/revitalization strategies 320, 326
see also Language endangerment; *specific
 languages*
- Eyak 252
see also Na-Dene languages
- Ezguerra, Domingo, Samar-Leyte 915
- ## F
- Facial expressions *see* Sign language
- Fali 3
see also Adamawa-Ubangi languages
- False friends *see* Borrowing
- Fanagalo 411–412
 classification 249t
 influence from other languages
 Afrikaans 411
 English 412
 Nguni 412
 Xhosa 411
 Zulu 412
 origin/development 411
 structure 412

- Fanagalo (*continued*)
 tense markers 412
 use of 411
 decrease 411
 verb inflexions 412
 word order 412
see also Bantu languages; Creoles; Pidgins;
 Xhosa; Zulu
- Fanakalo, SVO 412
- Fante dialect, Akan 17
- Faroe Islands, Danish 279
- Ferguson, Charles A, Ethiopian linguistic area (ELA) 379
- Fiji 413
 Fijian 412, 413
- Fijian **412–413**
 classification 250–251
 grammar 413
 literary tradition 413
 as official language, Fiji 413
 phonemes 413
 use of 412
see also Austronesian languages; Oceanic languages
- Finisterre-Huon languages 669, 670*f*
- Finite verbs
 in agglutinating languages 416*t*
 in introflecting language 51, 51*t*
- Finland 911
- Finnic languages
 classification 1129–1130
 consonant gradation 1130
 diphthongs 1130
 subordinate sentences 1132
 vowel harmony 1130
see also Uralic languages
- Finnish (Suomi) **413–415**
 adjectives 414
 as agglutinating language **415–420**
 aspect 1131–1132
 case marking 415
 classification 1129–1130
 consonants 413–414, 1130
 dictionary 415
 diphthongs 414
 fusion 419
 long-range comparisons
 Amerind 655
 sound correspondences with Pipil 651
 morphology 414, 415
 mutation 418
 nominals 416, 416*t*, 417*t*
 case forms 416
 nouns 414
 numerals 414
 object marking 1132
 orthography 414
 perfect 414
 phonology 413
 plural markers 1131
 polyfunctional suffixes 419
 possessives 416*t*
 pronouns 414
 stem morphophonological alternations 417
 allomorphy 417
 alternations 418–419
 assibilation 418–419
 consonant gradation 418
 vowel harmony 417–418
 vowel mutation 418
 weak grade 418, 418*t*
 stress 414
 suffix morphophonological
 alternations 419
 allomorphs 419
 consonant deletion 419
 syntax 415
 use of 413
 verb inflectional class 417*t*
 verbs
 finite verb forms 414, 416*t*
 main verb phrases 1132
 nonfinite verb forms 414–415, 416*t*
 verb inflectional class 417
- vowel harmony 414, 417–418
- vowels 413–414
- word order 1132
 SVO 413
 VSO 413
- word stress 1131
- word structure 415
see also Arabic, as introflecting
 language; Central Siberian Yupik;
 Finnic languages; Morphological Types;
 Polysynthetic languages
- Finno-Ugric languages 254
see also Uralic languages
- First Sound Shift (Grimm's Law) 448*t*
- Fleming, H C, language families 652–653
- Flemish *see* Dutch
- Florentine *see* Italian
- Flores languages **420–421**
 classification 250–251, 420
 consonants 420
 metaphor 420
 symbolism 420
 verbal morphology 420
 vowels 420
see also Austronesian languages;
 Central Malayo-Polynesian (CMP);
 Malayo-Polynesian languages
- Focus-Mood-Aspect morphology, Bikol 159*t*
- Focus particles 989
- Folk characters, North American native language
 variation 756
- Folklore/folktales, North American native
 language variation 759
- Fon (Fon-gbe) 631–632
see also Gbe languages
- Fopo 624
see also Grebo languages
- For, use of 775*f*
- Foreign sign languages 954
- Forest Nenets 761–762
see also Nenets (Yurak)
- Formal languages *see* Artificial languages
- Formosan languages **421–425**
 classification 250–251, 421, 422*f*
 dictionaries 422–423
 grammars (books) 422–423
 history 422
 noun phrase 423
 research history 422
 Bible translation 422
 structure 423
 subgrouping 422
 verbal affixes 423–424
see also Amis; Austronesian languages;
 Ayatalic languages; Malayo-Polynesian
 languages
- Four-way stop contrast, Tanoan languages
 1049–1050
- Frafra 472
see also Oti-Volta languages
- France
 Basque 144–145
 Breton 166
 Catalan 188
 Dutch 307
 French 427
 German 444
 Occitan 799
- Franglais **425–427**, 429
 damage to French language 425–426
 definition 425
 determiner use 425
 development 425
 bilingualism 425
 contact 425
 humor 425
 English words 425
 Étiemble, René 425
 French words 425
 graphemes 425
 Kington, Miles 425
 orthography effects 425
 phonology effects 425
see also French
- Frankish, influence on other languages, French
 429
- Francisco León
 derivational nouns 714
 phonology 713
see also Mixe-Zoquean languages
- Fraser, John, Pictish 856
- French **427–430**
 adjectives 429
 as analytic language 428–429
 augmentative 428–429
 classification 251–252
 damage to language,
 Franglais 425–426
 determiners 429
 dialects/dialectology 427
 diminutives 428–429
 as fusional language 428
 future 428
 influence from other languages
 English *see* Franglais
 Old Norse 248
 influence on other languages
 Bislama 162
 Dutch 309
 Early Modern English 341
 English 248
 Inuit 374
 Jèrriais 563
 Korean 615–616
 Krio 620
 Lesser Antilles 858–859
 Michif 710
 Middle English 352, 354
 Modern Standard Arabic (MSA) 43
 Old English 358
 Sango 917
 Wolof 1186–1186
 interrogation 429
 long-range comparisons
 chance similarities 652
 sound correspondences 650–651
 morphology 428
 negation 429
 as official language
 Canada 427
 Central African Republic 917
 France 427
 French Polynesia 1039
 Madagascar 674
 Vanuatu 161
 origin/development 427
 orthography 428
 acute accents 428
 cedilla 428
 circumflex 428
 grave accents 428
 Roman alphabet 428
 tréma 428
 perfect 428
 phonetics 427
 Canada 427–428
 consonants 428
 liaison 428
 open syllables 428
 vowels 427
 phonology 427
 subjects 429
 SVO 429
 syntax 428
 tenses 428
 use of 427
 vocabulary 429
 Arabic 429
 English 429
 Franglais 425
 Frankish 429
 Italian 429
 Latin 429
see also Franglais; Jèrriais; Middle English;
 Romance languages
- French-English cognates 651
- French Guiana, Arawak languages 59
- French Polynesia, official language 1039

- Frisian
 classification 251–252
 Dutch, influences from 310
see also Germanic languages
- Frisian, Old, Old English *vs.* 356–357
- Friulian
 classification 893
 dialects 894
 geographical distribution 894
 speaker numbers 894
see also Rhaeto-Romance languages
- Front/back harmony, vowels *see* Vowel harmony
- Fujian 969
see also Hakka languages
- Fula 770
 classification 253
 Hausa, influences on 477
 language families 652–653
 use of 770
see also Atlantic Congo languages
- Fulfulde 430–433
 concord 432
 consonant alteration 432*t*
 dialects 431
 diminutives 431
 focus 432
 voices 432
 Islam 430
 nomenclature 430
 noun classes 430, 431
 suffix allomorphy 431
 orthography 430
 Arabic script 430
 Latin script 430
 taboos 432
 body parts 432
 names 433
 use of 430
 as L2 430
see also Atlantic Congo languages
- Fullah, Krio, influences on 620
- Fusion, in agglutinating languages 419
- Fusional languages 291, 731, 732
 agglutinating languages *vs.* 550
 affixes *vs.*, roots 554
 classification vagueness 549
 definition 549
 index of fusion 291
 index of synthesis 291
 isolating languages *vs.* 550
 morphological types 549
 other types *vs.* 733*t*
see also specific languages
- Future
- African-American Vernacular English (AAVE)
 335–336
- Bactrian 116
- Balkan linguistic area *see* Balkan linguistic area
- Bantu languages 141
- Bikol 159*t*
- Bislama 162
- Brahui 164*t*, 165
- Bulgarian 169
- Creek 265
- Creole Portuguese 868
- Creoles 863
- Czech 277
- Dravidian languages 304
- Evenki 406
- French 428
- Gamilaraay 440–441
- Gondi 457
- Hausa 479
- Hiligaynon 494
- Ilocano 521
- Italian 552–553
- Jiwarli 572*t*
- Kannada 576
- Krio 619
- Kurukh 628*t*
- Lak 636
- Latin 643
- Lithuanian 648
- Luganda 658
- Lyngngam 596
- Malayalam 683
- Mixe-Zoquean languages 712
- negation 129*t*
- Nenets (Yurak) 762–763
- Nivkh 778
- Oneida 807
- Ossetic 817*t*
- Persian, Modern 851
- Pitjantjatjara 873*t*
- Pitta Pitta 88
- Punjabi 888
- Russian 907
- Sindhi 963*t*
- Siriano 1098*t*
- Slovak 979
- Tajik Persian 1043
- Tibetan 1061
- Tohono O'odham 1075
- Tok Pisin 1077
- Tucanoan languages 1101, 1101*t*
- Turkish 1114–1115
- Waray-Waray 915–916
- will/have* 128, 129*t*
- Wolaitta 1182
- Xhosa 1192
- Yiddish 1204
- Fuzhou 219
 phonology 219
 Putonghua *vs.* 219
 use of 219
see also Chinese
- G**
- Ga-Dangme 631
 double articulated consonants 632
 tone 632
 verbs 632
 vowel harmony 632
 word order 632
see also Kwa languages
- Gaelic, Scots *see* Scots Gaelic
- Gagauz 1109
 related languages, Azerbaijanian 110–111
 use of 1112
see also Turkic languages
- Galician 435–438
 allophones 437
 classification 251–252, 435
 definite articles 437
 history 435
 morphology 436
 phonology 435, 436*t*
 Portuguese, influence on 883
 Portuguese *vs.* 435
 pronouns 437
 syntax 437
 tenses 436–437
 use of, Spain 435
 verbs 437
see also Portuguese; Romance languages
- Gallatin, Albert, Uto-Aztecan languages 1140
- Gambia
 Mande 769–770
 Wolof 1184
- Gamilaraay 438–441
 Bible translation 438
 case 439
 consonants 439, 439*t*
 future 440–441
 history 438
 morphology 439
 nouns 439
 phonology 439
 pronouns 440, 440*t*
 relationships 439
 revival/survival 438–439
 roots 439
- stops 439
 study 438
 syntax 440
 interclausal syntax 440–441
 use of 438
 verbs 440
 conjugations 440*t*
 dependent verbs 440
 vowels 439
 word-building 440
 word order 440
see also Australian languages; Jiwarli; Pama-Nyungan languages
- Gan languages
 classification 969
 speaker numbers 214*t*
see also Chinese
- Garbrialiño 1140
see also Uto-Aztecan languages
- Garo 968–969
see also Bodo-Koch languages
- Gartner, T, Rhaeto-Romance language
 classification 893
- Gascon *see* Occitan
- Gatschet, Albert, Uto-Aztecan languages 1140
- Gaulish
 Breton *vs.* 166
 classification 199–200
see also Celtic, Continental
- Gawarbat
 case-marking 284
 classification 282
 speaker numbers 282
 tonal system 526–527
see also Indo-Aryan languages; Kunar languages
- Gbaya languages 3
 long-range comparisons 651
see also Adamawa-Ubangi languages
- Gbe languages 631–632
 double articulated consonants 632
 vowel harmony 632
see also Aja (Aja-gbe); Kwa languages
- Gè 750
see also Native American languages
- Gedebo 624
see also Grebo languages
- Gedeo
 noun morphology 491*t*
 phonology 490*t*
 use of 488–489, 488*t*
 verb morphology 490, 490*t*
see also Highland East Cushitic (HEC) languages
- Geechee *see* Gullah
- Geelvink Bay languages
 geographical distribution 840–841
see also Papuan languages
- Ge'ez 441–442
 morphology 442
 phonology 441–442
 quasi-syllabic script 441–442
 use of 441
 vocabulary 442
 VSO 442
see also Afroasiatic languages; Amharic; Ethiopian linguistic area (ELA); Ethiopian Semitic languages; Semitic languages
- Gender
- Abun 1177
- Arabic 46, 55
- Arawak languages 61
- Bactrian 115, 540
- Bantu languages 140
- Barupu (Warupu) 974
- Berber languages 154
- Bilua 205
- Brahui 164
- Cariban languages 184–185, 186*t*
- Central Solomons languages 205
- Chorasmian 238–239, 540
- Creoles 862
- Cushitic languages 274–275
- Czech 277
- Dardic languages 284

- Gender (*continued*)
 Dutch 309
 English, Early Modern 340
 German 445–446
 Germanic languages 449
 Gondi 456
 Gujarati 527
 Ijò 518
 Indo-Aryan languages 527
 Iranian languages 540
 Italian 550–551, 551*t*
 Khoesaa languages 602
 Khotanese 540
 Kurdish 625–626
 Kurukh 627
 Latvian 645
 Lavukaleve 205
 Lithuanian 647–648
 Macedonian 664
 Makushi 184–185
 Manambu 693
 morphology 758
 North American native languages *see* North American native languages
 Omaha-Ponca 804
 Oneida 806–807
 Palikur 61
 Pemong 184–185
 Persian, Modern 540, 850
 Polish 875
 Punjabi 887
 Ramo 974
 Romani 899
 Savosavo 205
 Sindhi 962, 963*t*
 Sinhala 965, 966*t*
 Skou languages 974
 Slovak 978
 Slovene 983
 Sogdian 540
 Sumerian 1023
 Tajik Persian 1042
 Tariana 61, 1051
 Tocharian 1069–1070
 Touo (Baniata) 205
 Uralic languages 1131
 Wambaya 1163
 West Papuan languages 1177
 Wolaitta 1180, 1181–1182, 1181*t*, 1182*t*
 Yiddish 1204
 Yoruba 1208
- Genetic relationships, areal linguistics 66
 Genitive case, dative merging 125
 Genje 112
see also Azerbaijanian
- Georgia
 Armenian 68
 Caucasian languages 192
 Georgian 442
 Ossetic 812
- Georgian 442–444
 alphabet, Abkhaz 1
 classification 251
 consonants 443, 443*t*
 ergative 443
 Old 291
 phonetics, vowels 193, 193*t*
 use of 442
 verbs 443, 443*t*
 case marking/agreement 443*t*
 vowels 443, 443*t*
 written text 442–443
 oldest example 442
see also Abkhaz; Caucasian languages;
 Kartvelian languages
- German 444–447
 Central 445
 classification 251–252
 genders 445–446
 High 445
 history 445
 Bible translation 445
 High German 445
 Second Sound Shift 445
- written records 445
 inflectional marking 445–446
 influence on other languages 444
 Inuit 374
 Korean 615–616
 Pennsylvania Dutch 444
 Rabaul 444
 Slovak 980
 Tok Pisin 858–859
 Yiddish 444
- long-range comparisons
 chance similarities 652
 with English 651
- morphology 445
 orthography 445
 noun capitalization 445
- phonology 445
 regional/social variation 446
 related languages, Luxembourgish 659
 SOV 446
 syntax 446
 finite verbs 446
- Upper 445
 use of 444
- vowel inflection changes 446
see also Dutch; Europe, as linguistic area;
 Germanic languages; Luxembourgish;
 Sorbian
- Germanic languages
 classification 251–252
 genetic classification 246
 East 447
see also Gothic
 Esperanto, influence on 376
 gender distinctions 449
 migrations 448–449
 North 447
 nouns 449
 origin/development 447
 lexicon 447–448
 relationships between 448
 SOV 449
 Tocharian 1070
 West 447
 word order 449
see also Afrikaans; Danish; Dutch; English,
 Early Modern; English, Modern; German;
 Gothic; Icelandic; Indo-European
 languages; Luxembourgish; Middle
 English; Norse, Old; Norwegian; Old
 English; Swedish; Yiddish
- Germany
 Danish 279
 German 444
 Sorbian 991–993
 Urdu 1133
- Gerunds *see* Noun(s); Verb(s)
- Gesture *see* Sign language
- Ghana
 Akan 17
 Ewe 408
 Gur 770
 Gur languages 472
 Kwa 771
 Kwa languages 630
 Mande 769–770
 national languages, Ewe 408
- Gheg, Albanian dialects 23
- Gibraltar, Yanito 1202
- Gigimai 1074
see also Tohono O’odham
- Gikuyu (Kikuyu) 449–453
 aspect marking 452
 classification 253
 consonants 450
 diminutives 452
 diphthongs 450
 glides 450
 habitual phonology 450
 influences from other languages 449–450
 lexical tone 450
 nouns 450
 classes 450, 451*t*
 compound 452
- concord system 450
 derivation 452
 deverbal 452
 morphosyntax 450
 phonology 450
 prenasalized consonants 450
 SVO 451
 syllables 450
 tense marking 452
 as Thagicu language 449–450
 as tonal language 450
 triphthongs 450
 use of, Kenya 449–450
 verbs 451
 negation 452
 reduplication 451–452
 vowels 450
 word order 451
see also Bantu languages
- Gilchrist, John, Hindustani grammar 498
- Gilgiti 282
see also Gilgit languages
- Gilgit languages 282
see also Shina languages
- Gilij, Filippo Salvatore, Cariban language
 classification 183
- Gilligan, Gary, suffixing preference 288
- Gimira 805
see also Omotic languages
- Girard, V, Cariban language
 classification 183
- GiTonga 1018
see also Inhambane languages
- Glebo 624
see also Grebo languages
- Globalization, and language endangerment 325,
 326
- Global language, definition 326
- Glottochronology, Indo-European
 languages 529
- Goeje, C H de, Cariban language
 classification 183
- Goidelic Celtic
 classification 200, 251–252
see also Celtic; Celtic, Insular
- Goidelic languages 453–455
 definition 453
 Norse, effects of 453–454
 Ogham 453
 perfect 453
see also Celtic
- Gondi 455–459
 ablative 457
 adjectives 456
 adverbs 456
 agreement 300*t*
 case suffixes 457
 classification 251
 classifiers 457
 consonants 455, 455*t*
 dialects 455
 future 457
 gender 456
 genitive 457
 instrumental-locative suffix 457
 interjections 456
 nominative 457
 nouns 456
 accusative 301–302
 number 456
 numerals 457, 457*t*
 particles 456
 phonology 455
 plural suffixes 456
 postpositions 456, 457
 pronouns 457
 syntax 456
 use of 455
 India 455
 verbs 456, 457
 finite verbs 456, 457, 458*t*
 nonfinite verbs 457
 verb bases 457
 vowels 455, 455*t*

- word classes 456
word order 456
see also Dravidian languages
- Gonga 805
see also Omotic languages
- Gonja 631
see also Guang languages
- Gordon, Raymond G Jr., *Ethnologue* 385
- Gorokan languages
classification 1087
geographical distribution 669, 670f
grammars (books) 1085–1086
see also Trans New Guinea languages
- Gorum 736
morphology 737
see also Munda languages
- Gothic 459–461
classification 251–252
Crimean Gothic 460
history 459
influence on other languages 460
morphology 460
sample text 460
syntax 460
Wulfila 459
 alphabet 459, 459f
see also Germanic languages; Germanic languages, East; Indo-European languages; Old English
- Grammars (books)
Akkadian 21
Chadic languages 206
Formosan languages 422–423
Hindustani 497–498
Trans New Guinea languages 1085–1086
- Grammaticalization, Southeast Asian languages 1012
- Grammaticized verbs, evidentiality *see* Evidentiality
- Grangali
case-marking 284
classification 282
sibilants 283
speaker numbers 282
see also Kunar languages
- Grasserie, Raoul de la 833
- Grave, French orthography 428
- Great Britain *see* United Kingdom (UK)
- ‘Great Vowel Shift’
Early Modern English 339, 339t
Later Modern English 349
Scots 924–925
- Grebo languages 770
phonetics/phonology, syllables 624
use of 624
see also Kru languages
- Greece
Albanian 22
Macedonian 663
Modern Greek 464
Romanian 901
Turkish 1112
- Greek
Ancient *see* Greek, Ancient
Armenian *vs.* 68–69
classification 251
influence from other languages
 Phoenician 854
 Present Day English 329–330
influence on other languages
 English 248
 Yiddish 1205
Italy 545
modern *see* Greek, Modern
see also Indo-European languages
- Greek, Ancient 461–464
accent 462
adjectives 463
consonants 462
declensions 463
dialects 461
 ‘historical,’ 461
 ‘literary,’ 462
 Mycenaean 461
external history 461
Latin, relation to 461
Modern Greek *vs.* 464–465
morphology 462
nominals 463
orthography, Mycenaean 461
perfect 463
phonology 462
resultative 463
stops 462
syntax 462
tenses 463
types 463
verbs 463
vocalic resonants 462
vowels 462
word order 463
see also Balkan linguistic area; Greek, Modern; Latin
- Greek, Modern 464–467
Ancient Greek *vs.* 464–465
dialects 465
 Peloponnesian-Ionian Greek 465
 Pontic 465
 Tsakonian 465
as fusional language 466
as official language, Greece 464
origin/development 464
perfect 466
sociolinguistic setting 465
stress accents 466
structure 465
syntax 466
as synthetic language 466
use of 464
vowels 465
see also Balkan linguistic area; Greek, Ancient
- Greek script, Bactrian 115
- Greenberg, Joseph H
African languages 4, 248–249
 Adamawa-Ubangi languages 2
 Benue-Congo languages 150
 Gur language studies 472
 Khoasaan language classification 600–601
 Kordofanian languages 613
 Kru language classification 623
 Mande language classification 696–697
 Niger-Congo languages 769f
 Nilo-Saharan languages 4, 772
American languages 248–249, 252
Amerind 655
 Macro-Jê language classification 665–666
 Songai language classification 991
morphological types 730, 731
multilateral comparison 650
Oceanian languages
 Central Solomon language classification 204–205
 Papuan languages 839–840
- Greenland
Danish 279
Greenlandic (Kalaallisut) 1172
- Greenlandic (Kalaallisut) 1172
Central Siberian Yupik *vs.* 202
classification 1172
dictionary 1172
grammar 1172
historiography 1172
use of 1172
workers in 1172
- Greenlandic, East 1172
see also Greenlandic (Kalaallisut)
- Grimes, Barbara G, *Ethnologue* 385
- Grimes, Joseph E, fields of work, three-letter language identifiers 386
- Grimm’s Law, French-English cognates 651
- Grundriss der vergleichenden Grammatik der indogermanischen Sprachen* (Brugmann) 135
- Grusi languages 770
see also Gur languages
- Gruzinic *see* Judeo-Georgian
- Guajiro 40
see also Arawak languages
- Guambiano *see* Barbaconan languages
- Guangdong 969
see also Hakka languages
- Guang languages 631
see also Awutu; Kwa languages
- Guangxi 969
see also Pinghua languages
- Guarani 467–468, 752
active/stative verbs 467
case markers 468
diminutives 467
history 467
initial consonants 468
morphology 467
as official language 467
as polysynthetic language 467
prosodic nasality (nasal harmony) 468
relations 468
use of 467
see also Native American languages; Tupian languages; Tupi-Guarani
- Guarekena (Warekena) 60
see also Arawak languages
- Guatemala 653
Arawak languages 59
Chalchiteko 705–706
Mayan languages 705–706, 706t
official languages 705–706
- Guató
classification 665–666, 666t, 667
inflectional morphology 667
phonology, vowels 667
word order 667–668
see also Macro-Jê languages
- Guaycuruan 752
see also Native American languages
- Guere-Krahn 624
see also Guere languages
- Guere languages 770
types 624
use of 624
 speaker numbers 623
see also Kru languages
- Guernsey, languages, French 427
- Guinea
Fulfulde 430
Mande 769–770
Mande languages 694
- Guinea-Bissau
Mande 769–770
Portuguese 883
- Gujarati 468–470
classification 251–252, 522
dialects 468
genders 527
grammar 469
history 468
literature 468
number of speakers 523
as official language 468
orthography 469
phonology 525–526
use of 468
vowels 526–527
see also Dardic languages; Indo-Aryan languages; Indo-Iranian languages
- Gulf Zoque *see* Mixe-Zoquean languages
- Gullah 470–472
classification 249t
development 470
etymology 470
habitual 471
history 470
influence from other languages 471
iterative 471
lexicon 471
phonology 471
pronominal form 471
use of 470
see also African-American Vernacular English (AAVE); Creoles; English; Pidgins
- Gun (Gun-gbe) 631–632
see also Gbe languages
- Gundert, Hermann, Malayalam 681

- Gunwinygu (Gunwigu)
 morphology 89–90
 semantics, pronominal forms 90–91, 91*t*
see also Australian languages
- Guosa, as auxiliary language 75–76
- Guoyu *see* Putonghua
- Gurage languages 383
 use of 382–383
see also Ethiopian Semitic languages
- Gurjuc *see* Judeo-Georgian
- Gur (Voltaic) languages 770
 aspect 473
 classification 768–769
 Central Gur 472
 Senufo 472
 concord 473
 consonants 472
 grammar 472
 imperfective 473
 phonetics/phonology 472
 plurals 473
 studies of 472
 Polyglotta Africana 472
 subgroups 770
 SVO 473
 syntax 472
 tones 473
 use of 770
 Mali 472
 speaker numbers 472
 vowel harmony 472
 word order 473
 workers in 472
see also Adamawa-Ubangi languages; Niger-Congo languages
- Gurmukhi script, Punjabi writing
 systems 886
- Gutob 736
see also Munda languages
- Guugu Yimithirr 473–476
 cardinal directions 475
 downstep 473
 genetic relations 474
 Kuku Yalanji 474
 history 473–474
 orthography 474
 personal pronouns 474
 progressive 474–475
 shortening *vs.* lengthening suffixes 474
 sociolinguistic features 476
 suffixes 474
 syntax 475
 use of 473–474, 476
 verbs 474
 transitive verbs 475
 vowels 474
 word order 474
 workers in, Roth, W E 473–474
see also Australian languages
- Guyana
 Arawak languages 59
 Hindi 495
- Gypsy *see* Romani
- ## H
- Haarmann, H, European linguistic area 391
- Habitual
 African-American Vernacular English (AAVE) 336
 Bengali 149
 Crow 269
 Dravidian languages 304–305
 Gikuyu (Kikuyu) 450
 Gullah 471
 Iranian languages 539
 Iroquoian languages 544
 Krio 619
 Ossetic 816, 817*t*
 Papiamentu 835
 Pidgins 867–868
 Sindhi 963*t*
 Sumerian 1024
- Hadiyya
 noun morphology 491*t*
 phonology 490–491, 490*t*
 use of 488–489, 488*t*
 Ethiopia 272–273
see also Cushitic languages; Highland East Cushitic (HEC) languages
- Hadramitic 931
see also Semitic languages
- Hadza 252
see also Khoesaa languages
- Haisla 1157
see also Wakashan languages
- Haitian Creole French, Louisiana Creole *vs.* 656
- Haketiya 567
- Hakka languages
 classification 969
 speaker numbers 214*t*
see also Chinese; *specific languages*
- Halang
 speaker numbers 726
see also Bahnaric languages
- Hale, Kenneth L
 Arrernte study 73
 Warlpiri morphology 1165
- Hamar 805
see also Omotic languages
- Hamitic theory, Afroasiatic languages 12
- Hamito-Semitic languages *see* Afroasiatic languages
- Handshape, phonology 941, 941*f*
- Han'gul, Korean alphabet 616
- Hani 968–969
see also Lolo-Burmese languages
- Hanzi 217
- Harmony systems *see* Consonant(s); Vowel harmony
- Hasaitic 931–932
see also Semitic languages
- Haspelmath, Martin, European linguistic area 393
- Hatam
 classification 1176
 numbers 1177
see also East Bird's Head (EBH) languages; West Papuan languages
- Hattic, kinship 196
- Hausa 206, 477–480
 consonants 477, 477*t*
 future 479
 ideophones 479
 imperfective 478
 influence from other languages 477
 Krio, influences on 620
 morphology 478
 negation 479
 nouns 478
 phrases 479
 phonology 477
 plurals 478
 possessives 479
 pro-drop 479
 pronouns 478
 questions 479
 reduplication 478
 subject agreement 478
 SVO 478
 syntax 479
 tense-aspect/mood 478
 tones 478
 use of 477
 media 477
 Nigeria 477, 1209
 verbs 478
 vowels 477
 word order 479
 written script 477
see also Africa; Africa, as linguistic area; Afroasiatic languages; Chadic languages
- Have perfect tense
 Balkan linguistic area 129
 Standard Average European (SAE) languages 393–394
- Hawaii, Cebuano 197
- Hawaiian 480–481
 first recording 1149–1150
 Hawaiian Creole English, influences on 481
 morphophonemics 1150
 phonemes 1150
 revival of 1149
see also Hawaiian Creole English (HCE); Tahitian
- Hawaiian Creole English (HCE) 481–482
 classification 249*t*
 graded variation 481
 grammar 481
 influence from other languages 481
 literary traditions 481–482
 phonology 481–482
 SVO 481
 use of
 speaker numbers 481
 USA 1127
 VSO 480–481
see also Creoles; Hawaiian; Pidgins
- Hawking, John A, suffixing
 preference 288
- Head driven Phrase Structure Grammar (HPSG), Balinese 118
- Hebraica *Stuttgartensia* 483
- Hebrew
 classification 250
 influence on other languages
 Jewish languages 566
 Judeo-Arabic 568
 Yanito 1202
 Yiddish 567, 1205
 Israeli *see below*
 Middle 934
 Phoenician, influence from 854
 Phoenician *vs.* 854
 pre-Modern *see below*
 Rabbinic 483
 the Torah 483–484
 Yiddish alphabet 1203
see also Afroasiatic languages; Aramaic; Semitic languages
- Hebrew, Israeli 485–488
 adjectives 487
 as analytic language 486
 clauses 487
 consonants 486–487
 genetic classification 485
 grammatical profile 486
 as head-marking language 486
 influence on other languages, Jewish languages 566
 nouns 487
 origins/development 485, 486*f*
 phonology 486
 political influences 488
 pronouns 487
 syllable structure 487
 use of, Israel 485
 verbs 487
 vowels 486–487
see also Afroasiatic languages; Hebrew; Jewish languages; Semitic languages; Yiddish
- Hebrew, pre-Modern 482–485, 933
 Biblical 482
see also Afroasiatic languages; Jewish languages; Semitic languages
- decline of 483
 definition 482
 in Diaspora 483
 literature 484
 secular context 484
 Israeli pronunciation 484
 Masoretes 483
 Middle 934
 in other languages 484
 Rabbinic 483
 as 'sacred language,' 482

- study in Christianity 484
in Talmud 483
Torah 483–484
- Heiban 770
see also Kordofanian
- Heiltsuk 1157
see also Wakashan languages
- Henry, George, Nyanja 794
- Henshaw, H W, Native American languages 747
- Heritage languages
definition 318
endangerment *see* Language endangerment
- Heterograms, Pahlavi (Middle Persian) 827
- Hetherwick, Alexander, Nyanja 794
- Hibito-Cholón languages 41
see also Andean languages
- Hidatsa 749
see also Siouan languages
- High German 445
- Highland East Cushitic (HEC) languages 272, 488–492
classification 489, 489f
demography 488
members 488–489, 488t
noun morphology 491, 491t
phonology 490, 490t
progressive 490
SOV 489
syntax 491, 491t
types 489, 489t
use of 272–273, 488–489
Ethiopia 272–273, 488–489
verb morphology 490, 490t
writing 489
see also Africa; Africa, as linguistic area;
Afroasiatic languages; Cushitic languages;
Ethiopian linguistic area (ELA)
- Hildegard of Bingen, Saint, artificial languages 76
- Hiligaynon 492–494
consonants 493
deictics 493t
dialects 492
future 494
glottal stop 493
grammar 493
history 492
morphology 493
negatives 494
nouns 493
phonology 492t, 493
progressive 493t
pronouns 492t
use of 99, 492
Philippines 492
speaker numbers 492
verb inflection 493–494, 493t
vowels 493
word order 494
see also Austronesian languages; Bikol; Samar-Leyte; Tagalog
- Hill (western) Mari 1129–1130
see also Mari languages
- Hindi 494–497
Assamese *vs.* 78
causatives 997
classification 251–252, 522
compounding 497
consonants 495–496
converbs 995
dative subjects 997
derivation 497
dialects 495
ergative 496–497
formal *vs.* informal 495
influence on other languages
Hindustani 497
Kashmiri 582–583
Kurukh 629
Malayalam 680–681
Punjabi 889
influences from other languages 495–496
learning by children 495
morphology 496–497
- Nepali *vs.* 764
nouns 496–497
as official language 495
Fiji 413
India 499
origin/development 494–495
phonology 495–496, 496t
political influences 495
postposition 496–497
reduplication 497
SOV 497
tenses 497
- Urdu *vs.* 1134
lexicography 500
use of 495
number of speakers 522–523
verbs 496–497
vowels 495–496, 526–527
word order 497
writing systems, Devanagari 524
see also Bengali; Dardic languages; Hindustani;
Indo-Aryan languages; Indo-European
languages; Punjabi; Romani; South Asian
languages; Urdu
- Hindko 635
- Hinduism
Indo-Iranian languages 531–532
language influences, Khmer
(Cambodian) 600
Malayalam 680
Punjabi 886
revivalism, Hindustani effects 499
see also Vedas
- Hindustani 497–501
classification 251–252
dictionaries 500
forms of 498
bazaar 499
vernacular 499
grammars (books) 497–498
Hindu revivalism 499
influences from other languages 497
lexicon 500
Perso-Arabic words 500
Muslim revivalism 499
origin/development 497
colonial influences 498
dialect mixing 498
Kari Boli 497
literary language 498
problems of linguistic description 500
religious influences
Islam 498
partitioning 499
as symbol of unity 499
Urdu, divergence from 1135–1136
vocabulary 499
see also Dardic languages; Hindi; Indo-Aryan
languages; Urdu
- Hiri Motu 501–502
consonants 501
definition 501
OSP 501
postposition 501
sentence structure 501
SOP 501
vowel sounds 501
- Hishkaryana
geography 185f
phonology 183–184
word order 186
see also Cariban languages
- Hismaic 932
see also Semitic languages
- Hispano-Celtic *see* Celtiberian
- Historicist studies, linguistic areas 62
- Hitchiti 739
- Hitchiti-Mikasuki 749
see also Muskogean languages
- Hitrite 502–503
cuneiform script 502
example 502
historical aspects 36
Luwian *vs.* 37
- morphology 502
phonology 502
syntax 502
texts 532
see also Indo-European languages
- Hmar 968–969
see also Kuki-Chin languages
- Hmong
classifiers 1013
come to have verb 1015
word order 1014t
- Hmongic 503
see also Hmong-Mien (Miao-Yao)
languages
- Hmong-Mien (Miao-Yao)
classification 968
ideophones 503
see also Sino-Tibetan languages
- Hmong-Mien (Miao-Yao) languages 503
numbers speaking 503
used in 503
see also Hmongic
- Ho 736
see also Munda languages
- Hoenigswald, H M, long-range comparisons 650
- Hokan languages 750
adjectives 508
alignment 507
classification 505
classifier 508
comparative studies 504
consonants 507t
contrasts 507t
core additions 504
descriptive works 506
diminutives 507–508
ergative 507
grammar 507
Hokan hypothesis 504
interrogatives 508
members 504
noun phrases 509
nouns 507
Oto-Mangean *vs.* 505
person markers 507
phonemic contrasts 506
phonology 506
Pomo family 750–751
possessives 509
quantifiers 508
sentence-level constituents 509
SOV 507
syllable structure 507
use of 505
verbs 508
viability 509
vowels 507t
word order 509
workers in 504
Yuman family 750–751
see also Achuan languages; Chimariko
languages; Washo; Yanan languages
- Hokan-Siouan languages 747–748
- Honduras
Arawak languages 59
Misumalpan languages 711
Sumu (Sumo Tawahka) 711
- Hong Kong Cantonese 218
consonants 219
derivation 218
English, influences from 218
phonology 220t
speaker numbers 218
syllables 219
use of 218
vowels 219
see also Chinese
- Honorifics
Japanese 559
Korean 615
Telugu 1056, 1056t
Tibetan 1062
- Ho Nte 503
see also Hmong-Mien (Miao-Yao) languages

- Hope-Tewa 1049
see also Tanoan
 Hopi 1140
 baby talk vocabulary 513–514
 classification 1139*t*
 demonstratives 512
 development 511
 dialects 513
 imperfective 512
 male *vs.* female speakers 513
 nouns 512
 plurals 512
 orthography 511
 pronouns 512
 ritual speech 514
 song 514
 source information types 513
 SOV 511
 time 511
 use of 511
 verbs 512
 derivational suffixes 512–513
 modifiers 513
 noun incorporation 513
 tense 512
 transitive verbs 512
 word order 511
 workers in, Whorf, B L 511
see also Keres; Uto-Aztecan languages
 HPSG (Head driven Phrase Structure Grammar),
 Balinese 118
 Hre 726
see also Bahnaric languages
 Hua 1085–1086
see also Gorokan languages
 Huave 748
see also Native American languages
 Huayu *see* Putonghua
 Hübschmann, Heinrich, Armenian 68–69
 Huehuetla Tepehua 1081
 phonology, consonants 1082
 speaker numbers 1082
 Hü:hu'ula 1074
see also Tohono O'odham
 Huhuwos 1074
see also Tohono O'odham
 Hui languages
 classification 969
 speaker numbers 214*t*
see also Anhui; Chinese
 Huilliche 41
 classification 701
see also Andean languages
 Huli 1086–1087
see also Engan languages
 Humahuaca 41
 Human pronouns, in isolating language 221
 Humboldt, Wilhelm von, morphological types
 730
 Humburi Senni 990–991
see also Songai languages
 Humor, Franglais development 425
 Hung 728–729
see also Cuoi languages
 Hungarian 514–516
 case suffixes 1131
 classification 1129–1130
 consonants 515
 definite objects 515
 history 514
 language names 514
 main verb phrases 1132
 morphology 515
 negation 1131
 nouns 515
 noun phrase 516
 object marking 1132
 phonology 514
 plural markers 1131
 possessed nominals 515
 postpositions 290, 515
 pro-drop 516
 syntax 515
 use of 514
 verbs 515
 vowel harmony 1130
 vowels 514–515
 vowel harmony 514–515
 word order 515
 word stress 515
 written records, earliest 514
see also Australian languages; Finno-Ugric
 languages; Uralic languages
 Hungary
 German 444
 Hungarian 514
 Slovak 977
 Slovene 981
 Hunzib
 vowels 193, 194*t*
see also Caucasian languages
 Huon-Finistere languages 1086–1087
see also Trans New Guinea languages
 Huron 542
see also Iroquoian languages
 Hurrian 516
 as agglutinating language 516
 concord 515–516
 extinction 516–516
 historical aspects 516
 nouns 516
 proper names 516
 Hutton, James, Indo-European languages
 528–529
 Huzhu Mongghul 723
see also Mongol languages
- I**
 Ibaloi *see* Inibaloi (Ibaloi)
 Ibanag 784
see also North Philippine languages
 Ibo 620
 Icelandic
 classification 251–252
 neologisms 782
 patronyms 782
 speaker numbers, absolute 323
see also Danish; Germanic languages; Indo-
 European languages; Middle English;
 Norse, Old; Norwegian; Scandinavian
 languages; Swedish
 Icelandic, Old 779–783
 diphthongs 780
 'eddic' poetry 779–780
 historical relations 779
 indefinite article 780
 morphology 780
 negation 781
 nominative case 781
 noun phrases 781
 numbers 780
 phonology 780
 'scaldic' poetry 779–780
 syntax 780
 terminology 779
 use of, Norway 779
 verb-initial clauses 781
 verbs 780–781
 vocabulary 781
 vowels 780
 word order 781
 Ideophones
 Aweti 1106–1107
 Chadic languages 207
 Ethiopian linguistic area (ELA) 379
 Ethiopian Semitic languages 383
 Ewe 408
 Hausa 479
 Hmong-Mien (Miao-Yao) 503
 Kamayurá 1106–1107
 Karitiana 1106–1107
 Karo 1106–1107
 Kinyarwanda 608
 Mawé 1106–1107
 Mondé 1106–1107
 Mundurukú 1106–1107
 Ramaráma 1106–1107
 Shona languages 938
 sign language 946
 Tupari 1106–1107
 Tupian languages 1106–1107
 Wolof 1184–1185
 Xhosa *see* Xhosa
 Xipáya 1106–1107
 Ido 76–77
 Idomoid 151
see also Benue-Congo languages
 Igbo
 classification 253
 Nigeria 1209
see also Benue-Congo languages
 Igboid 151
see also Benue-Congo languages
 Ijò 517–518
 classification 517
 Defaka *vs.* 517
 as Kwa language 517
 west *vs.* east 517
 dialects 518
 downstep 518
 gender 518
 geographical location 517
 Nigeria 517
 naming of 517
 nomenclature 517
 noun class 518
 qualifiers 518
 SOV 517–518
 typological characteristics 517
see also Kwa languages
 Ijoid 770
 classification 768–769
see also Niger-Congo languages
 Ika *see* Chibchan languages
 Ikpeng
 phonology 183–184
 use of 185*f*
see also Cariban languages
 Illich-Svutych, Vladimir M
 long-range comparisons 653–654
 Nostratic theory 249
 Ilocano 518–522
 consonants 518–519, 519*t*
 demonstratives 521, 521*t*
 dialects 518
 future 521
 noun marking system 520, 520*t*
 potentive mode 520
 pronouns 520, 521*t*
 stops 519
 stress 519
 syntax 520
 use of 99
 Philippines 518
 speaker numbers 783
 verbs 520
 voices 520, 520*t*
 vowels 519, 519*t*
see also Austronesian languages; North
 Philippine languages; South Philippine
 languages
 Ilonggo *see* Hiligaynon
 Imhambane languages *see* Bantu languages,
 Southern
 Imperfective
 Arabic 47
 Berber languages 156
 Bulgarian 168
 Central Semitic languages 931
 Cupeño 271
 Cushitic languages 274–275
 Czech 277
 Dogon 294
 Ethiopian linguistic area (ELA) 380
 Ethiopian Semitic languages 383
 Gur languages 473
 Hausa 478
 Hopi 512
 Kalkutungu 575
 Macedonian 664–665

- Nenets (Yurak) 762–763
 Ossetic 816
 Papiamentu 835
 Pashto 847
 Phoenician 855
 pidgins 867–868
 Pitjantjatjara 873
 Polish 876
 Russian 907
 Slovak 977
 Slovene 983
 Sorbian 994
 Tohono O'odham 1075
 Totonacan languages 1083
 Warlpiri 1166*t*
 Wolaitta 1182
 Xhosa 1192
 Inari Saami 911
 see also Saami
 Incorporation
 Cherokee 544
 Cree *see* Cree
 Crow 269
 Hopi 513
 Iroquoian languages 544
 Karajá 667
 Lakota 639
 Mohawk 544
 Nuuchahnulth (Nootka) 790
 Panárá 667
 Indefinite articles, Standard Average European (SAE) languages 393–394
 Indeterminateness, Southeast Asian languages 1011
 Index of fusion, types 291
 Index of synthesis
 classification of language 731
 types 291
 India
 Bengali 148
 Burmese 170
 Burushaski 175
 Dardic languages 282
 Dhivehi 285
 Dravidian 297
 Gondi 455
 Gujarati 468
 Hindi 495, 499
 Indo-Aryan languages 522
 Indo-Iranian languages 531
 Kannada 576
 Kashmiri 582
 Khasi 595
 Khasic 727
 Kurukh 626–627
 Malayalam 680
 Marathi 703
 Mon-Khmer languages 725
 Munda languages 736
 Nepali 764
 Punjabi 885
 Santali 921
 Sindhi 960
 Sino-Tibetan 968
 Syriac 1033
 Tai-Kadai languages 105
 Tai languages 1039
 Telugu 1055
 Tibetan 1060–1061
 Urdu 522–523, 1133
 Indo-Aryan languages 522–528
 case inventory 527
 classification 251–252
 contacts
 Dravidian languages 306
 Munda languages 737
 distribution 522
 English, influences from 522
 ergative 527
 genders 527
 Indo-Iranian languages *vs.* 525
 major languages 522
 morphosyntax 527
 origins/development 523
 literary traditions 524
 Middle Indo-Aryan dialects 523
 Old Indo-Aryan dialects 523
 written records 523
 phonology 525
 sociolinguistics 522
 stops 525–526
 tense system 527
 tonal system 526–527
 use of 522
 number of speakers 522–523
 vowels 526–527
 writing systems 524
 Brahmi 524
 Devanagari 524
 Kharosthi 524
 see also Asoka; Assamese; Italian; Kashmiri; Lahnda; Marathi; Nepali; Punjabi; Romani; Sanskrit; Sindhi; South Asian languages; Southeast Asian languages; Tocharian; Urdu
 Indo-European languages 528–531
 classification 251, 530
 archaeology 530
 convergence 530
 genetic classification 246
 nostraticism 530
 Nostratic theory 249
 types 530
 definition 530
 family tree construction 528
 Hutton, James 528–529
 'Neogrammarians,' 528–529
 Schleicher, August 528–529
 as fusional languages 550
 historical language change theories 529
 'glottochronology,' 529
 Schmidt, Johannes 529
 and Nostratic 653–654
 Nostratic theory 786
 Proto-Indo-European *vs.*
 cladistics 529
 Nichols, Johanna 529
 Tocharian *vs.* 1070
 workers in
 Hutton, James 528–529
 Jones, William 528
 Nichols, Johanna 529
 Schleicher, August 528–529
 Schmidt, Johannes 529
 Young, Thomas 528
 see also Afrikaans; Albanian; Anatolian languages; Armenian languages; Avestan; Catalan; Germanic languages; Goidelic languages; Gothic; Hindi; Hittite; Icelandic; Indo-Iranian languages; Italian; Italic languages; Latin; Norse; Old; Nuristani languages; Old Church Slavonic; Persian, Old; Pictish; Proto-Indo European (PIE); Slavic languages; South Asian languages; Spanish; Tocharian
 Indo-Iranian languages 531–535
 active *vs.* passive 534
 allophones 533
 Armenian *vs.* 68–69
 classification 251–252
 demonstrative pronouns 533
 Hittite texts 532
 Indo-Aryan languages *vs.* 525
 inflection 533
 labiovelars 532
 lexicon 534
 moods 534
 nominal compounds 534
 nouns 533
 origin/development 531–532
 perfect 533–534
 perfect tense 533–534
 phonology 532
 present tense 533
 pronouns 533
 religious influences 531–532
 SOV 534
 syntax 534
 tense/aspect distinctions 533
 use of 531
 verbs 533
 vowels 532
 word order 534
 see also Avestan; Dardic languages; Gujarati; Indo-Aryan languages; Indo-European languages; Iranian Languages; Iranian languages; Kashmiri; Pashto; Persian, Old; Sanskrit
 Indonesia
 Austronesian languages 97, 99
 Balinese 116
 Cebuano 197
 Madurese 672
 Malay 678
 Indonesian (Bahasa Indonesia)
 Malay 678
 Riau Indonesian *vs.* 895–896
 Indus Khosistani
 sibilants 283
 speaker numbers 282
 Iñeri 59
 see also Arawak languages
 Infixes, diachronic origins 287
 Inflection
 genetic classification 246
 nonnative English 361
 sign language morphology 950, 951*f*
 Inflectional phrase (IP) *see* Sentence
 Ingáin 666–667
 see also Jê languages
 Ingrian 1129–1130
 see also Finnic languages
 Inhambane languages 1017
 Inheritance, sign language 940
 Iniai (Bisorio) 1086–1087
 see also Engan languages
 Inibaloi (Ibaloi) 784
 see also North Philippine languages
 Ink-brush styles, Chinese script *see* Chinese script
 Inner Mongolian 969
 see also Jin languages
 Innes, Gordon, Kru language studies 623
 Innovation spread, language diffusion 246–248
 Inscriptions
 Etruscan 388
 Italic languages 555
 Insular Celtic *see* Celtic, Insular
 Intensifier-reflexive differentiation, Standard Average European (SAE) languages 393–394
 Intergenerational variation, North American native languages 755
 Interior Salish 749
 see also Salishan languages
 Interlingua 77
 International Auxiliary Language Association, artificial languages 77
 International Organization for Standardization (ISO), three-letter language identifiers 386
 International Phonetic Alphabet (IPA), Turkish consonants 1113, 1113*t*
 Introflexing languages, Arabic 50–53
 Intrusion effects, Italic languages 555
 Inuit 374
 dialects 374
 history 371
 influences from other languages 374
 use of 374
 see also Eskimo-Aleut
 Inupiaq 535–537
 classification 251
 consonants 536
 dialects 535
 ethnonyms 536
 influence from other languages 536
 lexicon 536
 morphology 536
 phonology 535
 as polysynthetic language 536
 SOV 536
 syntax 536
 use of 535
 geographical distribution 535, 535*f*

- Inupiaq (*continued*)
 USA 535
 viability 536
 verbs 536
 vowels 536
 word order 536
 workers in 535–536
 writing 535–536
see also Eskimo-Aleut languages; West Greenlandic
- Invented languages, literature *see* Artificial languages
- Ipili 1086–1087
see also Engan languages
- Iran
 Arabic 42
 Aramaic 56
 Armenian 68
 Azerbaijani 112–113, 1112
 Balochi 134, 538
 Brahui 162–163
 Iranian languages 537
 Kurdish 538, 625
 Modern Persian 538, 850
 Qashqay 1112
 Iranian languages 537–542
 classification 251–252
 consonants 540
 declensions 540, 541
 definite articles 540
 diphthongs 540
 directional prefixes 542
 direct object 541
 dual 540
 genders 540
 genetic relationships 538
 habitual 539
 historical 540
 imperfect tense 541
 influences on other languages
 Pahlavi (Middle Persian) 827
 Tocharian 1070
 Middle 537
 modal forms 542
 morphology 539
 New 538
 Arabic words 538
 dialects 538
 Turkish 538
 nominal systems 539
 nouns 541
 Old 537
 palatal affricates 538, 540
 past tenses 541
 perfect 537
 phonology 538
 plurals 540
 present tense 541
 pronominal systems 539
 pronouns 541
 punctual aspects 542
 script 537
 syllables 540
 syntax 539
 historical 540
 tenses 541
 use of 537
 verbs 539, 541
see also Aramaic; Avestan; Bactrian; Balochi; Chorasmian; Elamite; Indo-Aryan languages; Indo-Iranian languages; Iranian languages, Old; Khotanese; Kurdish; Ossetic; Pahlavi (Middle Persian); Pashto; Persian, Modern; Persian, Old; Romani; Sogdian; Tajik Persian; Tocharian; Turkic languages
- Iraq
 Iranian languages 537
 Kurdish 538, 625
- Irish
 classification 251–252
 development of 454
 Connacht 454
 Munster 454
 Ulster 454
 Eastern 454
 see also Scots Gaelic
 literary Irish 453–454
 sample 454
see also Goidelic Celtic; Goidelic languages
- Irish, Old 453
 active 453
 conjugated prepositions 453
 medio-passive 453
 nominals 453
 preterite 453
- Iron
 affricates 813
 classification 813
 consonants 814*t*
- Iroquoian languages 749
 agent prefixes 544
 classification 252
 consonants 543
 habitual 544
 laryngeal features 543
 local categories 544
 members 542
 morphology 543
 nouns 544–545
 incorporation 544
 particles 544–545
 perfect 544
 phonology 543
 pronominal prefixes 544
 stress 543
 subject/object categories 544
 syntax 543
 tone 543
 use of 542
 verbs 543–544, 544–545
see also Native American languages
- Irula
 Malayalam *vs.* 682
 nouns, accusative 301–302
see also Dravidian languages
- Ishtmus *see* Mixe-Zoquean languages
- IsiSwati *see* Swati
- Iskateko 819–821
see also Oto-Mangean languages
- Islam
 influence on
 Dhivehi 285
 Hindustani 498, 499
 languages
 Arabic 42
 Fulfulde 430
 Malayalam 680
 Punjabi 886
 Qur'an, Sindhi translations 961
 Urdu literature 1137
see also Arabic, Classical
- Island Melanesia, languages, Papuan languages 841
- Isoglosses
 European linguistic area *see* Europe, as linguistic area
 South Asian languages 1000
- Isolating languages 291, 731, 732
 definition 221
 fusional languages *vs.* 550
 index of fusion 291
 index of synthesis 291
 other types *vs.* 733*t*
see also specific languages
- Isopleth mapping, Africa, as linguistic area 6, 6*f*
- Israel
 Arabic 42, 485
 Armenian 68
 English 485
 Israeli Hebrew 485
 Israeli Hebrew *see* Hebrew, Israeli
 Israeli Sign Language, suffix allomorphs 942, 942*f*
- Istro-Romanian 901
- Italian 545–549
 adjectival inflection 550
 affixes 553
 homonymy 550
 inflection 550
 inflectional *vs.* derivational 553
 interfixes 553
 agreement 550
 classification 251–252
 cumulative exponence 550
 descendants 548
 dialects/dialectology 545–546
 example 548
 external history 546
 as fusional language 549–554
see also Arabic, as introflexing language; Central Siberian Yupik; Italian; Morphological Types; Morphological types; Polysynthetic languages
 genetic relationship 545
 individual characteristics 547
 influence on other languages 548
 French 429
 Yanito 1202
 influences from other languages,
 Latin 545–546
 internal history 546
 investigation history 548
 morphology 546, 547
 noun inflection 550
 gender 550–551
 as official language 545
 perfect 546
 phonology 546, 547
 rhythm 547–548
 pro-drop 547
 pronoun case traces 551
 clitic pronouns 551
 stressed pronouns 551
 sociolinguistic points 548
 dialect use 548
 literary use 548
 SVO 554
 syntax 546
 use of 545
 verbs 551
 analytic forms 552
 auxiliary verbs 552
 indicative future tense 552–553
 indicative imperfect 552
 indicative present conjugation 553
 irregular conjugations 553
 present condition tense 552–553
 preterit stem irregular
 modifications 553*t*
 suprasegmental modification 552
 thematic vowels 552
 word order mobility 554
 writing system 547
 alphabet 547
 written records 546
 earliest text 546–547
 vernacular 547
see also Indo-Aryan languages; Indo-European languages; Romance languages
- Italic languages 554–555
 classification 251–252
 genetic classification 246
 definition 554–555
 inscriptions 555
 intrusion effects 555
 reconstruction 246
 Umbrian language 555
see also Indo-European languages
- Italkic *see* Judeo-Italian
- Italy
 Albanian 545
 Catalan 188, 545
 French 427
 German 444, 545
 Greek 545
 Modern Greek 464
 Occitan 799
 official languages, Italian 545
 Serbo-Croatian 545
 Slovene 545, 981

- Itbayat
 phonology 684
 use of 783
see also Malayo-Polynesian languages; North Philippine languages
- Itelmen
 classification 239
see also Chukotko-Kamchatkan languages
 speaker numbers 239–240
- Iteration *see* Recursion/iteration
- Iterative
 Anatolian languages 37
 Cushitic languages 274–275
 Gullah 471
 Krio 617
 Mapudungan languages 702
 Polish 876
 sign language 941
 Slovak 979
 Sorbian 994
 Tagalog 1036–1037
 Tucanoan languages 1099–1100
 Warlpiri 1166
- Itzaj 705–706
 speaker numbers 706*t*
see also Mayan languages
- I-umlaut, Old English 357
- Ivatan 783
see also North Philippine languages
- Ivory Coast
 Eastern Kru languages 623–624
 Grebo languages 624
 Guere languages 624
 Gur languages 770
 Kru languages 770
 Kwa 771
 Kwa languages 630
 Mande languages 769–770
 Western Kru 624
- Ixcatec 751
see also Popolocan languages
- Ixil 705–706
 speaker numbers 706*t*
see also Mayan languages
- Ixilan 705–706
see also Mayan languages
- J**
- Jabo 624
see also Grebo languages
- Jabuti
 classification 665–666, 666*t*
 geographical distribution 666–667
 morphology 667
see also Macro-Jê languages
- Jackson, Kenneth Hurlstone, Pictish 856
- Jainism, Indo-Iranian languages 531–532
- Japan, Ainu 15
- Japanese 31
 as Altaic language 31
 and Amerind 655
 classification 249
 Nostratic theory 249
 development 557
 Altaic languages 557
 Austronesian languages 557
 geographical isolation 557
 dialects 557
 honorifics 559
 influence from other languages 557
 lexicon 557
 mimetic words 557–558
 literary history 557
 modifiers 558
 mora 558
 pronoun omission 559
 related languages 557
 Ainu 557
 Korean 31, 557
 Ryukyuan 557
 segmental phonology 558
 consonants 558
 vowels 558
 SOV 558
 speaker numbers 557
 syntax 558
 as tone language 558
 topic construction 558
 nontopic *vs.* 558
 word order 558
 writing system 557
see also Ainu; Altaic languages; Austronesian languages; Korean; Ryukyuan
- Japanese Sign Language (JSL) 956
- Japeria *see* Cariban languages
- Jaqaru 752
see also Aymaran
- Jaquaru languages, Aymara *vs.* 108
- Jaquari *see* Native American languages
- Java, Madurese 672
- Javanese 560–562
 Balinese, influence on 116–117
 consonants 560, 560*t*
 morphology 560
 phonology 560
 Sulawesi 560
 use of 99, 560
 vowels 560, 561*t*
 writing system 561
see also Austronesian languages; Balinese; Madurese; Malay
- Jeh
 speaker numbers 726
see also Bahnaric languages
- Jê languages
 classification 665, 666, 666*t*, 667
 ergative 668
 morphology 667
 as ergative language 668
see also Macro-Jê languages
- Jen languages 3
see also Adamawa-Ubangi languages
- Jenner, Edward, Cornish 258
- Jërriais 562–565
 affricates 562
 classification 251–252
 consonants 562
 delateralization 563
 dictionaries 564
 glides 563
 grammars (books) 564
 history 562
 influence from other languages 563
 language planning 564
 literary tradition 564
 morphosyntax 563
 phonology 562
 use of 562
 extracurricular teaching 564
 geographical distribution 562, 562*f*, 563*f*
 speaker numbers 562
 velar nasals 563
 vocabulary 563
 vowels 562
 workers in, Le Geyt, Matthieu 564
see also French; Romance languages
- Jersey 427
- Jewish English *see* English Jewish
- Jewish languages 120, 565–569
 biblical/liturgical translations 566
 common linguistic features 566
 definition 565
 history 565
 influence from other languages 566
 scholarship 565
 types 565
 use of 565
see also Afroasiatic languages; Hebrew, Israeli; Hebrew, pre-Modern, Biblical; Yiddish
- Jianghuai
 classification 214
 speaker numbers 214*t*
see also Mandarin
- Jiangxi 969
see also Hakka languages
- Jiaoliao
 classification 214
 speaker numbers 214*t*
see also Mandarin
- Jicaque 748
see also Native American languages
- Jidi *see* Judeo-Persian
- Jidyó *see* Judeo-Spanish
- Jin languages
 classification 969
 speaker numbers 214*t*
see also Chinese
- Jinuo 968–969
see also Lolo-Burmese languages
- Jirajaran *see* Arawak languages
- Jivaroan languages 41
see also Andean languages
- Jiwarli 569–574
 consonants 570, 570*t*
 demonstratives 571*t*
 example 573
 future 572*t*
 history 569–570
 language relationships 570
 Tharrkari 570
 Thiin 570
 Warriyanka 570
 morphology 570
 as non-configurational language 572–573
 nouns 570–571, 571*t*
 phonology 570
 pronouns 571, 571*t*
 root 570
 stops 570
 suffixes 570
 syntax 572
 use of 569–570
 verbs 571–572, 572*t*
 vowels 570
 word building 571
see also Australian languages; Gamilaraay; Pama-Nyungan languages
- Johnston, J B, Pictish 856
- Jomang 613
see also Kordofanian languages
- Jones, Charles, Later Modern English definition 343
- Jones, David, Malagasy writing systems 674–675
- Jones, William
 Indo-European languages 528
 Modern Persian 850
 Sanskrit 921
- Jordan, Domari 295
- Juang 736
see also Munda languages
- Judaism
 Aramaic scripture commentaries 58
 the Torah 483–484
see also Hebrew
- Judeo-Arabic 565
 development 568
 influence from other languages 568
 speaker numbers 568
 writings 568
- Judeo-Aramaic 565
 development 566
 use of 566
- Judeo-Berber 565
- Judeo-English 565
- Judeo-French 565
- Judeo-Georgian 565
- Judeo-German *see* Yiddish
- Judeo-Greek 565
 development 566
 use of 566
- Judeo-Italian 565
- Judeo-Malayalam 565
- Judeo-Persian 565
- Judeo-Portuguese 565
- Judeo-Spanish 565
 development 567
 Haketiya 567
 literary tradition 567
 use of 566
 speaker numbers 567

- Judeo-Tadjik *see* Judeo-Persian
 Judeo-Tat *see* Judeo-Persian
 Judezmo *see* Judeo-Spanish
 Jula 770
 see also Atlantic Congo languages
 Ju languages 601–602
 see also Khoesaa languages
 Jul'hoan languages 601–602
 see also Khoesaa languages
 Juquila *see* Mixe-Zoquean languages
 Juray 736
 see also Munda languages
 Jurúna
 classification 1106*t*
 tone system 1106
 see also Tupian languages
- K**
- Kaape Afrikaans 8, 9*f*
 Kachari 968–969
 see also Bodo-Koch languages
 Kachchi, related languages, Sindhi 961
 Kadugli languages 613
 see also Kordofanian languages
 Kafiri languages *see* Nuristani languages
 Kainantu languages 1087
 see also Trans New Guinea languages
 Kainji 151
 see also Benue-Congo languages
 KAKIBA 517
 Kako 140
 see also Bantu languages
 Kakutungu (Kalkutung)
 ergative 88
 morphology 88–89
 syntax 88–89
 see also Australian languages
 Kalaallisut *see* Greenlandic (Kalaallisut)
 Kalak 613
 see also Kordofanian languages
 Kala Lagaw Ya, Australia 79
 Kalam
 complex predicates 671
 nouns 671
 Kalam-Kobon languages
 predicates 1089
 verb root 1087
 see also Trans New Guinea languages
 Kalam Kohistani, morphosyntax,
 case-marking 284
 Kalanga 1017
 see also Shona languages
 Kalapuya 750
 see also Penutian languages
 Kalasha
 agreement patterns 284
 case-marking 284
 sibilants 283
 speaker numbers 282
 Kalenjin *see* Nilo-Saharan languages
 Kaleung 726–727
 see also Katuic languages
 Kalimantan, languages, Javanese 560
 Kalispel 749
 see also Salishan languages
 Kalkutungu 575–576
 antipassive 575
 applicative constructions 576
 core case marking 575
 dependent clauses 575–576
 ergative 575
 imperfective 575
 insubordination 576
 nouns 575
 perfect 575
 phonemes 575
 verbs 575
 see also Australian languages; Pama-Nyungan languages
 Kalmuk 723
 see also Mongol languages
- Kam 3
 see also Adamawa-Ubangi languages
 Kamakā
 classification 665, 666, 666*t*
 geographical distribution 666–667
 records of 667
 see also Macro-Jê languages
 Kamas 1129–1130
 see also Samoyed languages
 Kamayurá 1106–1107
 see also Tupian languages
 Kambaata
 noun morphology 491*t*
 phonology 490–491, 490*t*
 use of 488–489, 488*t*
 see also Highland East Cushitic (HEC) languages
 Kamrupi 78
 Kamsá *see* Barbacoan languages
 Kamviri 787
 see also Nuristani languages
 Kanakanavu
 classification 421
 research history 423
 see also Formosan languages
 Kannada 576–578
 classification 251
 compound verbs 996
 converbs 995
 future 576
 grammar 577
 history 576
 India 576
 language contacts 577
 linguistic theory 577
 literature 577
 Malayalam *vs.* 682
 nouns
 genitive 302
 plural suffixes 301
 numerals 303
 pronouns 302–303, 302*t*, 303*t*
 SOV 577
 structure 577
 tenses 304
 variation 577
 written scripts 297, 576–577
 see also Dravidian languages; Malayalam
 Kanuri 578–579
 classification 253
 consonant alterations 578
 downstep 578
 grammar 578
 Hausa, influence on 477
 script 578
 tone levels 578
 use of 578
 speaker numbers 772–773
 verbs 578
 see also Nilo-Saharan languages
 Kapampangan 579–581
 adjectives 579
 case marking 579
 determiners 579, 580*t*
 ergative 579, 580*t*
 existence 579
 focus constructions 580
 grammar 579
 lexical classes 579
 negation 579
 nouns 579
 phonology 579
 possession 579
 pronouns 579, 580*t*
 speaker numbers 783
 word order 579
 see also Malayo-Polynesian languages; North Philippine languages; Southeast Asian languages; Tagalog
 Kapong
 gender 184–185
 use of 185*f*
 vowels 183–184
 see also Cariban languages
- Kaqchikel 705–706
 nongenuine sound correspondences with English 651
 positionals 707
 speaker numbers 706*t*
 verbs 707*t*
 see also Mayan languages
 Karabagh 112
 see also Azerbaijani
 Karachay-Balkar 1109
 see also Turkic languages
 Karaim *see* Turkic languages
 Karajá
 classification 665, 666, 666*t*, 667
 inflectional morphology 667
 male *vs.* female speech 667, 667*t*
 noun incorporation 667
 phonology 667
 use of 666–667
 vowels 667
 see also Macro-Jê languages
 Karakalpak 588, 591
 related languages, Kazakh 589
 see also Kazakh
 Karakhanid 1110
 see also Turkic languages
 Kara-Kirghiz *see* Kirghiz
 Karamzin, N M, Russian 905
 Karanga 1017
 see also Shona languages
 Karankawa 751
 see also Native American languages
 Karao 784
 see also North Philippine languages
 Karelian 1129–1130
 see also Finnic languages
 Karelian *Sprachbund* 392–393
 Karen
 classification 253
 classifier 581
 see also Tibeto-Burman languages
 Karenic languages
 classification 968–969
 morphology 970
 see also Sino-Tibetan languages
 Karen languages 581–582
 classification 581
 number of speakers 581
 sample 581
 as Sino-Tibetan language 581
 SVO 581
 as Tibeto-Burman language 581
 typological characteristics 581
 tone system 581
 verb-medial sentence type 581
 use of 581
 Burma 581
 Thailand 581
 writing systems 581
 see also Kayah
 Karenni 968–969
 see also Karenic languages
 Kari Boli, Hindustani
 origin/development 497
 Karihona
 phonology 183–184
 use of 185*f*
 see also Cariban languages
 Karinya 185*f*
 see also Cariban languages
 Kariri (Kariri-Xocó)
 classification 665, 666, 666*t*
 ergative 668
 morphology 667
 records of 667
 use of 666–667
 word order 667–668
 see also Macro-Jê languages
 Kariri-Xocó *see* Kariri (Kariri-Xocó)
 Karitiána
 classification 1106*t*
 ideophones 1106–1107
 positional demonstratives 1106
 see also Tupian languages

- Karmali Santali 736
see also Munda languages
Karna:taka Saba:nusa:-sana 577
- Karo
 case marking 1106
 classification 1106*t*
 evidentiality 1108
 ideophones 1106–1107
 nouns 1106
 classification 1108
see also Tupian languages
- Kartvelian languages
 Abkhaz, influence on 2
 classification 251
 consonants 193
 kinship 196
 morphology 194
 Nostratic theory 249, 653–654, 786
 verbs 195
 word order 195
see also Caucasian languages
- Karuk languages 750–751
 classification 505
see also Hokan languages
- Kashmiri 582–585
 agreement patterns 284
 case-marking 284
 classification 251–252, 282, 522, 582
 complementation 584
 consonants 583
 correlative 583–584
 demonstrative pronouns 584
 dialects 582
 ergative 584
 example 584
 influences from other languages 582–583
 morphological causatives 997
 morphosyntax 583
 number of speakers 523
 origin/development 582
 phonology 525–526, 583
 postpositional language 583–584
 religious influences 582
 sibilants 283
 sociolinguistics 584
 split-ergative language 584
 SVO 583–584
 syllables 583
 use of 582
 India 582
 Pakistan 582
 speaker numbers 282
 verb phrases 584
 vocabulary 582–583
 vowel harmony 583
 vowels 526–527, 583
 word order 284, 583–584
 writing systems 282, 524
 Devanagari 583
 Sharada 583
see also Areal linguistics; Dardic languages; Indo-Aryan languages; Indo-Iranian languages; Kashmiri languages; Pakistan; Punjabi; Romani; Sindhi
- Kashmiri languages 282
see also Dardic languages
- Kashuyana
 phonology 183–184
 use of 185*f*
see also Cariban languages
- Kasimov Tatar 1052
- Kataang
 speaker numbers 726–727
see also Katuic languages
- Kati 787
 dialects 787
see also Nuristani languages
- Katla 770
see also Kordofanian
- Katu
 speaker numbers 726–727
 verbs 725
see also Katuic languages
- Katuic languages 724
 use of 726–727
see also Mon-Khmer languages
- Katupha's law, Bantu languages 139
- Katz, Dovid, Yiddish development 1205
- Kaufman, Terrence, Cariban language
 classification 183
- Kavalan 421
see also Formosan languages
- Kavirajamarga* 577
- Kawaki 752
see also Native American languages
- Kawésqar (Qawaskar) 41
 affiliations, Mapudungan languages 701
see also Andean languages
- Kayah 581
see also Karen languages
- Kayardild 585–586
 case inflexions 586
 case suffixes 585
 classification 250
 compass terms 586
 phonology 585
 possessives 585–586
 use of 585
 word lists 585
see also Australian languages; Language endangerment; Non-Pama-Nyungan languages; Tangkic languages
- Kaytetye 586–588
 kin nouns 587, 587*t*
 nouns 587
 personal pronouns 587, 587*t*
 phonology 586–587
 consonants 586–587, 586*t*
 resources 588
 use of 586
 speaker numbers 586
 verbs 587, 587*t*
 word stress 587
see also Arrernte; Australian languages; Pama-Nyungan languages; Warlpiri
- Kazak
 classification 112
see also Azerbaijanian
- Kazakh 588–591
 comparatives 590
 converbs 590
 dialects 590
 distinctive features 589
 evidentiality 590
 grammar 590
 language contacts 589
 lexicon 590
 loanwords 590
 as official language, Kazakhstan 588
 origin and history 588
 phonology 589
 consonants 589–590
 loanwords 590
 suffix vowels 589
 vowels 589
 plurals 590
 possessives 590
 present tense 590
 pronouns 590
 related languages 589
 Karakalpak 589
 Kipchak 589
 Noghay 589
 Uzbek 589
 use of 588
 Afghanistan 588
 China 588
 Kazakhstan 588
 Kyrgystan 588
 Mongolia 588
 Russia 588
 Tajikistan 588
 Turkmenistan 588
 Uzbekistan 588
 Xinjiang 1142
 written language 589
 Arabic script 589
- Cyrillic alphabet 589
 Roman alphabet 589
see also Altaic languages; Bashkir; Karakalpak; Kirghiz; Mongolia; Turkic languages; Uyghur; Uzbek
- Kazakhstan
 Kazakh 588
 official language, Kazakh 588
 Uyghur 1142
 Uzbek 1145
- Kazak-Kirghiz *see* Kazakh
- Kazan Tatar *see* Tatar
- Kazukuru 204
see also Central Solomons languages
- Kebu 631
see also Togo Mountain languages
- Kei 690
see also Malukan languages
- Kelo 772–773
see also Nilo-Saharan languages
- Kemiehua 729
see also Mon-Khmer languages
- Kentish, Old English dialect 356
- Kenya
 Cushitic languages 272–273
 Gikuyu 449–450
 Luo 658
 Swahili 1026
- Keres 591–593
 consonants 591–592
 history 591
 morphology 591–592
 numbers 591–592
 pronominal prefixes 591–592
 sociolinguistics 592
 Spanish, influence from 591
 vowels 591–592
see also Hopi
- Keresan 750
see also Native American languages
- Ket 593–595
 case system 593
 classification 249
 dialects 593
 as first language 593
 morphosyntax 594
 noun classes 593
 Russia 593
 tones 593
 verbs 593
see also Language endangerment
- Ketelaer, J. J., Hindustani grammar 497–498
- Kewa 1086–1087
see also Engan languages
- Khakas, related languages, Yakut 1200
- Khalaj 1109
see also Turkic languages
- Khalashi 526–527
see also Indo-Aryan languages
- Khalkha
 Altaic hypothesis 653
 use of 723
see also Mongol languages
- Khamti 1039
see also Tai languages
- Khanty
 case suffixes 1131
 classification 1129–1130
 main verb phrases 1132
 object marking 1132
 vocalism 1130
 vowel harmony 1130
 word stress 1131
see also Uralic languages
- Kharia 736
see also Munda languages
- Kharoshthi script, Indo-Aryan language writing systems 524
- Khasi languages 595–597, 724
 classification 250
 morphology 595–596
 noun classes 596
 SVO 595–596
 use of 595

- Khasi languages (*continued*)
 India 727
 verbs 725
see also Austroasiatic languages; Mon-Khmer languages
- Kha Tong Luan 728–729
see also Muong languages
- Kazakhstan
 German 444
 Kirghiz 610
 Russian 588
- Kherwarian 736
 morphology 737
see also Munda languages
- Khmer (Cambodian) 597–600, 727
 classification 250
 classifiers 599
 grammar 599
 influence from other languages 600
 Pāli 248
 Sanskrit 600
 Thai 600
 influence on other languages
 Lao 640
 Thai 1059
 as isolating language 599
 morphology 599
 noun phrases 600
 phonology/phonetics 598
 consonants 598, 598*t*
 diphthongs 598–599, 598*t*, 599*t*
 history 599
 syllables 598
 tonal contrasts 599
 unaspirated *vs.* aspirated stops 598
 vowels 598, 598*t*, 599*t*
 religious influences 600
 script 598
 use of 597
 verbs 600
come to have verb 1015
 directional verbs 1014
 serial verb constructions 725
 verb phrases 600
 vocabulary 600
 word classes 599
 word order 599–600, 1014, 1014*t*
 written records 598
see also Austroasiatic languages; Mon-Khmer languages
- Khmer languages 724, 727
see also Mon-Khmer languages
- Khmú 725
- Khmuic, Lametic languages, influence on 728
- Khmuic languages 724
 classification 727
 local variants 727
 use of 727
see also Mon-Khmer languages
- Khoekhoe
 Afrikaans, influence on 7–8, 9, 11
 click variants 602*t*
 tonal melodies 602, 602*t*
 use of 601
 Namibia 601, 602
see also Khoesaa languages
- Khoesaa languages 600–603
 classification 252, 600–601, 601*t*
 Bleek, Dorothea 600–601
 extinct varieties 601
 Greenberg, Joseph H 600–601
 Schulze, Leonhardt 601
 clicks 602
 development 602
 as endangered languages 601
 gender 602
 language shift 321
 morphology 602
 phonology 602
 click variants 602, 602*t*
 tonal melodies 602, 602*t*
 tonology 602
 SOV 602
 syntax 602
 use of 601–602, 601*f*
 word order 602
 workers in
 Bleek, Dorothea 600–601
 Greenberg, Joseph H 600–601
 Schulze, Leonhardt 601
see also Nilo-Saharan languages
- Khotanese 537, 603–604
 cases 604
 classification 251–252
 directional prefixes 542
 genders 540
 lexicon 604
 modal forms 542
 perfect 604
 phonology 603
 intervocalic voiced stops 603–604
 palatal affricates 540
 retroflex consonants 603
 voiced sibilants 603
 SVO 604
 tenses 604
 potentialis tense 604
 Tocharian, influence on 1070
 verbs 604
 written records 603
see also Iranian languages
- Khowar
 agreement patterns 284
 Burushaski, influence on 179
 case-marking 284
 sibilants 283
 speaker numbers 282
 tonal system 526–527
 writing systems 524
see also Indo-Aryan languages
- Khün 1039
see also Tai languages
- Khusro, Amir, Hindustani 497–498
- K'iche' 705–706
 long-range comparisons 653
 positionals 707
 speaker numbers 706*t*
see also Mayan languages
- K'iche'an 705–706
 noun classifiers 708
see also Mayan languages
- Kickapoo 26
see also Algonquin languages
- Kikongo, influence on other languages,
 Palenquero 829
- Kikuyu *see* Gikuyu (Kikuyu)
- Kildin Saami 911
see also Saami
- Kilham, Hannah, Yoruba 1207
- Kim 3
see also Adamawa-Ubangi languages
- Kington, Miles, Français 425
- Kinship expressions
 Tohono O'odham 1075
 Warlpiri 1168
- Kinyarwanda 604–610
 adjectives 608
 classification 253
 class markers 607, 607*t*
 consonant phonology 607
 cluster consonants 605
 complex consonants 605
 consecutive consonants 607
 palatalized consonants 605
 prenasalized consonants 605
 prevelarized consonants 605
 simple consonants 605
 existential construction 609
 ideophones 608
 morphology 607
 multiple direct objects 609
 nouns 607
 object–subject reversal 609
 orthography 605
 phonology 606
 consonants *see above*
 Dahl's law 607
 gliding 607
 open syllables 605
 palatalized fricatives 607
 reduplications 607
 vowels *see below*
 word stems 605
see also Below, tonology
- possessives 608
 prefix derivation 608
 preprefix 608
 questions 609
 relative pronouns 609
 SVO 607
 syntax 609
 tense-aspect-modality 608
 tonology 606
 function 606
 lexical tones 606
 morphological tones 606
 nouns 606
 rules 606
 syntactic tones 606
 tense-mood-aspect morphemes 606
 verbs 606
 use of 604
 verbs 608
 serial verb construction 609
 vowel harmony 607
 vowel phonology 605, 606
 vowel coalescence 606
 vowel harmony 605
 word order 609
see also Bantu languages
- Kiowa 750
see also Tanoan
- Kiowa-Tanoan languages 750
- Kipchak, related languages, Kazakh 589
- Kipeá 667
see also Karirí (Karirí-Xocó)
- Kiranti languages
 classification 969
see also Sino-Tibetan languages
- Kirghiz 610–613
 dialects 612
 distinctive features 611
 grammar 612
 language contacts 611
 lexicon 612
 loanwords 611–612
 as official language, Kyrgyzstan 610
 origin and history 610
 phonology 611
 labialization 611
 morphophonemes 611
 vowels 611
 plurals 612
 possessives 612
 pronouns 612
 related languages 611
 Southern Altay Turkic *vs.* 611
 suffixes 612
 use of 610
 written language 611
 Arabic script 611
 Cyrillic alphabet 611
 Roman alphabet 611
see also Altaic languages; Kazakh; Turkic languages; Uyghur
- Kiriaka 841
see also North Bougainville languages
- Kisi-Sherbro 770
see also Atlantic Congo languages
- 'Kitchen Kaffir' *see* Fanagalo
- Kitsai *see* Caddoan languages
- Klamath-Modoc 750
see also Penutian languages
- Klao 770
 Liberia 624
see also Kru languages
- Klingon 76
- Ko:adk 1074
see also Tohono O'odham
- Koasati (Coushatta) 613, 738–739
 agreement type 741
 vowel length 740

- word order 741–742
see also Kordofanian languages
- Kobon
 complex predicates 671
 nouns 671
- Koch 968–969
see also Bodo-Koch languages
- Kochi 526–527
see also Indo-Aryan languages
- Kochimi languages 506
see also Hoka languages
- Kodagu
 Malayalam *vs.* 682
 nouns
 accusative 301–302
 genitive 302
 personal suffixes 305
see also Dravidian languages
- Koeber, Alfred Louis, Hoka languages 504
- Koelle, Sigismund Wilhelm
 Gur language studies 472
 Kru language studies 623
 Mande language classification 696
 Niger-Congo languages 768
 Yoruba 1207
- Kogui *see* Chibchan languages
- Kohistani languages 282
see also Dardic languages
- Koho 726
see also Bahnaric languages
- Koiari languages 1087
see also Trans New Guinea languages
- Kol 841
see also Munda languages; Papuan languages
- Kolarian *see* Munda languages
- Koló:di 1074
see also Tohono O'odham
- Kolyma *see* Yukaghir
- Koman
 classification 773
 consonants 774
 use of
 distribution 775f
 speaker numbers 772–773
see also Nilo-Saharan languages
- Komi
 classification 1129–1130
 word order 1132
see also Permian (Permian) languages
- Komi-Permyak 1129–1130
see also Permian (Permian) languages
- Komi-Zyrian
 case suffixes 1131
 classification 1129–1130
see also Permian (Permian) languages
- Konkani
 classification 522
 number of speakers 523
 phonology 525–526
 vowels 526–527
see also Indo-Aryan languages
- Konkomba 472
see also Oti-Volta languages
- Konobo 624
see also Guere languages
- Konyak 968–969
see also Konyak languages
- Konyak languages 968–969
see also Sino-Tibetan languages
- K'op 707t
see also Mayan languages
- Kopitar, Jernej, Slovene grammar 981
- Korafe 1085–1086
see also Binandere languages
- Koraku 736
see also Munda languages
- Koran, Sindhi translations 961
- Kordofanian languages 770
 branches/individual languages 613
 classification 768–769
 Greenberg, Joseph H 613
 noun classes 613
 subgroups 770
- use of 770
 speaker numbers 613
see also Adamawa-Ubangi languages; Niger-Congo languages; Nyanja
- Korea, Sino-Tibetan languages 968
- Korean 614–617
 classification 249
 as agglutinative language 615
 as Altaic languages 31
 Nostratic theory 249
 classifiers 615
 grammar 615
 honorifics 615
 influence from other languages 615–616
 link to Japanese 31
 markers 615
 number 615
 onomatopoeia 615–616
 phonology 614
 consonantal position 614
 initial consonant nonclustering 614
 intermorphemic sound change 614
 intermorphemic sound movement 614
 triple consonantal structure 614
 triple vowel system 614
 vowel stress 614
 postposition 616
 pronouns 615
 related languages 614
 Japanese 557
 sentence linkage 615
 speech level 615
 syntax 615
 vocabulary 615
 lexicon 615
 parallel vocabulary sets 616
 word creation 616
 written language 616
 word order 615
 word relation 615
see also Altaic languages; Japanese
- Koreguaje
 consonants 1095
 evidentiality 1100
 nasalization 1095
 noun classifiers 1098
 speaker numbers 1092t
 syllable pattern 1095
 verbs, evidentiality 1100
 word order 1096
see also Tucanoan languages
- Korekore 1017
see also Shona languages
- Korku 736
see also Munda languages
- Korwa 736
see also Munda languages
- Koryak
 classification 239
 speaker numbers 239–240
see also Chukotko-Kamchatkan languages
- Kosovo
 Albanian 22
 Macedonian 663
- Kota, Malayalam *vs.* 682
- Koyong 726
see also Bahnaric languages
- Koyra Chiini
 classification 990–991
 tone 774
see also Songai languages
- Kpelle 697–698
see also Mande languages
- Kposo 631
see also Togo Mountain languages
- Krachi 631
see also Guang languages
- Krahn 624
- Krenák (Botocudo)
 classification 665, 666, 666t, 667
 morphology 667
 use of 666–667
see also Macro-Jê languages
- Krio 617–623
 Bible translation 618
 classification 249t
 as tone language 622
 derivation 618
 development 617
 Domestic hypothesis 617
 Jamaican hypothesis 617
 future 619
 grammar 619
 habitual 619
 influence from other languages 620
 English 617
 French 620
 Niger-Congo languages 617
 Portuguese 620
 Spanish 620
 Temne 618
 iterative 617
 lexical features 620
 compounds 620
 reduplication 621
 literature 618
 morphemes 619
 negation 619
 perfect 619
 phonology 621
 consonants 621
 diphthongs 622
 suprasegmentals 622
 vowels 622
 plurals 619
 possessives 617, 619
 progressive 619
 serial verbs 619
 tenses 619
 use of
 in education 618–619
 as official language 618
 Sierra Leone 617, 618
 speaker numbers 618
see also Afrikaans; Creoles; Pidgins
- Krobu 631
see also Tano languages
- Krongo 613
see also Kadugli languages
- Kru languages 770
 classification 768–769
 Eastern Kru 623–624
 Greenberg, Joseph H 623
 concord 624
 dictionary 623
 grammar 624
 Krio, influence on 620
 noun classes 624
 perfect 624
 phonetics/phonology 624
 register tone 624
 syllables 624
 studies 623
 subgroups 770
 SVO 624
 syntax 624
 use of 770
 vowel harmony 624
 word order 624
 workers in 623
see also Aizi; Kwa languages; Niger-Congo languages
- Kuanhua 729
see also Mon-Khmer languages
- Kufo 613
see also Kadugli languages
- Kuikuro
 person-marking prefixes 185–186
 phonology 183–184
- Kuiper, Franciscus Bernardus Jacobus,
 fields of work, South Asian languages
 998–999
- Kuki-Chin languages
 classification 968–969
see also Anal; Asho Chin; Chin (Tiddim; Tedim); Sino-Tibetan languages; Tedim (Tiddim; Chin)

- Kuku Yalanji, genetic relations, Guugu Yimithirr 474
- Kuliak
classification 773
use of, distribution 775*f*
see also Nilo-Saharan languages
- Kulon 421
see also Formosan languages
- Kuman (Chimbu)
classification 1086–1087
contacts, Tatar 1053
pronouns 1087–1089
see also Chimbu-Wahgi languages
- Kumanakoto 185*f*
see also Cariban languages
- Kumauni
tonal system 525–526
vowels 526–527
see also Indo-Aryan languages
- Kumyk 1109
see also Turkic languages
- Kunama
use of 775*f*
vowel harmony 773–774
see also Nilo-Saharan languages
- Kunar languages 282
see also Dardic languages
- Kunua (Rapoiisi) 841
see also North Bougainville languages
- Kuot 841
see also Papuan languages
- Kurdish 625–626
case distinctions 625–626
classification 251–252
derivational nominal suffixes 626
dialects 625–626
dictionaries 625
genders 625–626
grammars (books) 625
influence from other languages 625
literary types 625
passive forms 626
past tenses 625–626
phonology 625
pronouns 625–626
use of 538, 625
verbs 626
see also Iranian languages; Persian, Modern
- Kurdistan, languages, Aramaic 58
- Kurdi *see* Judeo-Aramaic
- Kurschat, F, Lithuanian grammar 648
- Kuruáya 1106*t*
see also Tupian languages
- Kurukh 626–630
adjectives 627
adverbs 627
agreement 627, 628*t*
classification 251
future 628*t*
gender 627
Hindi, influence from 629
nouns 628
ablative 628
accusative 628
case suffixes 628
dative 628
genitive 628
instrumental 628
nominative 628
numerals 629
plural suffixes 628
postpositions 628
pronouns 628
number 627
phonology 627
consonants 627, 627*t*
vowels 627, 627*t*
postpositions 628, 629
syntax 627
use of 626–627
India 626–627
speaker numbers 626–627
verbs 629
agent noun 630
compound verbs 996–997
female speech 629
finite verbs 628*t*, 629
negative verb 630
nonfinite verbs 629
verb bases 629
word classes 627
word order 627
see also Dravidian languages
- Kurumba 301–302
see also Dravidian languages
- Kurux-Malto 299, 300*t*
see also Dravidian languages
- Kusal 472
see also Oti-Volta languages
- Kutenai 751
see also Native American languages
- Kutkashen 112
see also Azerbaijani
- Kuwaa 624
see also Kru languages
- Kuy
causative verbs 725
speaker numbers 726–727
see also Katuic languages
- Kwahu dialect, Akan 17
- Kwakiutlan branch, Wakashan languages 749–750
- Kwakwala 1157
classifiers 1160*t*
diminutives 1159*t*
possession 1160
suffixes 1159*t*
syllable structure 1158
see also Wakashan languages
- Kwa languages 771
classification 630, 768–769
concord 632
double articulated consonants 632
downstep 632
nouns 632
progressive 632
subgroups 771
Lagoon languages 631
SVO 632
tone 632
use of 771
verbs 632
vowel harmony 632
word order 632
workers in 630
Yoruba *vs.* 631
see also Abbey; Abé; Abidji; Adamawa-Ubangi languages; Adjukru; Akan; Alladian; Attié; Avikam
- Kwalean languages 1087
see also Trans New Guinea languages
- Kwikateko 819–821
see also Oto-Mangean languages
- Kwomtari languages 840–841
see also Papuan languages
- Kyrgystan
Kazakh 588
Kirghiz 610
official language 610
Uzbek 1145
- L**
- Lachik (Lasi) 968–969
see also Lolo-Burmese languages
- Ladin
classification 893
dialects 894
geographical distribution 893–894
speaker numbers 894
written records 893–894
variants 893–894
see also Rhaeto-Romance languages
- Ladino *see* Judeo-Spanish
- Lahnda 635–636
classification 251–252, 522
phonology 525–526
tonal system 526–527
see also Dardic languages; Indo-Aryan languages; Kashmiri; Pashto; Punjabi
- Lahu
classification 968–969
morphology 970
syllable structure 969–970
see also Lolo-Burmese languages
- Lak 636–638
deictics 636
future 636
locational affixes 636
morphology 636
noun classes 636
phonology 636
consonants 636, 636*t*
vowels 636
syntax 636
use of 636
Russia 636
word order 636–637
see also Caucasian languages; Daghestanian languages; North Caucasian languages
- Lake Miwok 651, 653
- Lakota 638–639
classification 252
dialects 754, 755, 755*t*
diminutives 755
men's speech *vs.* women's speech 639
morphology 638
nouns 638
noun incorporation 639
postpositions 638
syntax 639
use of 638
speaker numbers 638, 971–972
verbs 638
word classes 638
word derivation 638
affixation 638
circumstantial stems 638
compounding 638
word order 639
see also Caddoan languages; Central Siberian Yupik; Crow; Polysynthetic languages; Siouan languages
- Lala 1018
see also Nguni languages
- 'La langue d'oc' *see* Occitan
- Lamaholot 420
see also Flores languages
- Lamani 526–527
see also Indo-Aryan languages
- Lamet 728
classification 727
see also Lametic languages
- Lametic languages, Khmuic, influence from 728
- Langenhoven, C J 9–10
- Language
attrition *see* Language endangerment
bilingualism *see* Bilingualism
classification *see* Classification (of languages)
death *see* Language endangerment
definition 385
as continua 385
geographic distribution 319, 319*t*
literacy and *see* Writing/written language
maintenance *see* Language endangerment
sign *see* Sign language
Language and Culture Atlas of Ashkenazic Jewry 1206
- Language endangerment 317–327, 386
assessment 322, 386
criteria 322
education and literacy materials 324
governmental/institutional attitudes 325
intergenerational transmission 322
response to new domains and media 324
speaker numbers, absolute 323
total population speaker proportion 323
trends in existing domains 323
causes 325
definition 317
effects 320

- identification 387
 levels 319
 reversal/revitalization strategies 320, 326
 taxonomy of situations 321
see also Caddoan languages;
 Chukotko-Kamchatkan languages;
 Crow; Disappearing languages;
 Ethnologue; Kayardild; Ket; Language
 endangerment; Language shift; Louisiana
 Creole; Malukan languages; Mongolia;
 Tamambo; Uralic languages; Zapotecan
 languages
 Language prestige, Balkan linguistic area 132,
 132f
 Language shift
 causes 325
 definition 318, 319
see also Language endangerment
 Languages of wider communication *see*
 Bilingualism; World Englishes
 Language spread, English 363
 Language subgroupings, areal linguistics 65, 65t
 Languedocian *see* Occitan
 Lanyin
 classification 214
 speaker numbers 214t
see also Mandarin
 Lao 639–640
 classification 253
 classifiers 639–640
 grammar 639
 influence from other languages 640
 nouns 639–640
 phonology 639
 recent history 640
 sample of 640
 SVO 639–640
 Thai *vs.* 639
 as tonal language 639–640
 use of 639
 verbs 639–640
see also Tai-Kadai (Zhuang-Dong) languages;
 Tai languages
 Laos
 Bahnaric languages 725–726
 Burmese 170
 Katuic languages 726–727
 Khmuic languages 727
 Lao 639
 Mon-Khmer languages 725
 Palaung-Wa languages 727
 Tai-Kadai languages 105
 Tai languages 1039
 Viet-Muong languages 728–729
 Larteh 631
see also Guang languages
 Laru 613
see also Kordofanian languages
 Late Aramaic 934
 Late Egyptian 39
 Latin 640–644
 adjectives 642
 affiliations 640
 Ancient Greek *vs.* 461
 classification 251–252
 ‘deponents,’ 642–643
 derivational morphology 643
 future 643
 history 640
 inflection 642
 influence on other languages
 Catalan 190
 Dutch 309
 Early Modern English 341
 English 248
 French 429
 Italian 545–546
 Middle English 352, 354
 Old English 358
 Old Icelandic 781
 Portuguese 883
 Present Day English 329–330
 Romanian 900–901
 Slovak 980
 Spanish 1020
 Yiddish 1205
 language shift 321
 morphology 642
 nouns 642
 perfect 642
 phonology 641
 accent 642
 consonants 641t
 prosody 642
 semivowels 641
 spelling relationships 641
 vowels 641
 script 641
 alphabet 641
 sentence structure 643
 syntax 643
 varieties 641
 vulgar Latin 641
 written 641
 verbs 642, 643
 Wackernagel’s Law 643
 word order 643
see also Balkan linguistic area; Greek, Ancient;
 Indo-European languages; Italic languages;
 Middle English; Romance languages;
 Spanish
 Latin (Roman) alphabet
 Azerbaijani 111
 Balinese 117
 Dhivehi 286
 Evenki 405–406
 Fulfulde 430
 Kazakh 589
 Kirghiz 611
 Old English 357
 Polish 874–875
 Slovak 977–978
 Turkmen 1117
 Uyghur 1143
 Uzbek 1146
 Yoruba orthography 1207
 Latin America, Spanish 1020
 Latino sine Flexione 77
 Latvia, official languages 644
 Latvian 644–646
 classification 251–252
 consonants 645
 dictionaries 644
 gender 645
 grammars (books) 644
 influences 645
 lexicon 645
 nouns 645
 as official language 644
 origin/development 644
 stress 645
 use of 644
 verbs 645
 vowels 644–645
 workers in 644
 written works 644
see also Baltic languages; Balto-Slavic
 languages; Lithuania; Lithuanian
 Latviesu valodas vardnica 644
 Laumbe *see* Lavukaleve
 Laven (Bolovent) 726
see also Bahnaric languages
 Lavukaleve 841
 classification 204
 gender 205
 use of 204
see also Central Solomons languages
 Lawngwaw (Maru) 968–969
see also Lolo-Burmese languages
 Learning, by children 495
 Lebanon, Domari 295
 Leco 41
see also Andean languages
 Lectal level, pidgins 867
 Lectal variation, Creoles 867
 Left May languages 840–841
see also Papuan languages
 Le Geyt, Matthieu 564
 Leko languages 3
see also Adamawa-Ubangi languages
 Lelemi-Lefana 631
see also Togo Mountain languages
 Lembena 1086–1087
see also Engan languages
 Lenition, Wakashan languages 1158, 1159t
 Lepontic 199–200
see also Celtic, Continental
 Lepsius, Karl Richard, Hamitic theory 12–13
 Leslau, W, Ethiopian linguistic area (ELA)
 378–379
 Lesotho
 official languages 1017–1018
 Southern Bantu languages 1017
 Southern Sotho 1017–1018
 Lesser Antilles
 French, influence from 858–859
 lexicon 858–859
Lettsche grammatik 644
 Lëtzebuergesch *see* Luxembourgish
 Lewotobi 420
see also Flores languages
 Lewy, E 392
 Lexical categories, postbases 203
 Lexical comparison 650
 Lexical Functional Grammar (LFG), Balinese 118
 Lexical semantics
 Creoles 866
 nonnative English 360–361
 Lexical terms
 Afroasiatic languages 14
 Nostratic theory 786
 Lexicase *see* Case
 Lexicon *see* Word(s)
 Lexicostatistics, Tai-Kadai (Zhuang-Dong)
 language classification 248
 Lhuyd, Edward, Cornish 258
 Liberia
 Bassa 624
 Grebo languages 624
 Klao 624
 Kru 770
 Mande 769–770
 Mande languages 694
 Western Kru 624
 Libico-Berber script *see* Berber
 Libya, Berber 152–153
 Libyan script *see* Berber
 Lichtenstein, German 444
 Ligurian *see* Italian
 Lilloet 749
 Limba, Krio, influence on 620
 Limousin *see* Occitan
The Lindesfarne Gospels 358
 Lingala 137
see also Bantu languages
 Lingua franca, definition 318
 Lingua Ignota 76
 Linguistic areas
 borrowings *vs.* 67
 circumstantial studies 62
 definition 3, 62, 64
 examples 62
 historicist studies 62
 types 66
see also Areal linguistics; *specific linguistic
 areas*
A Linguistic Atlas of Late Middle English 355
 Linguistic communities, World Englishes 364
 Linguistic creativity, multilingualism 368
 Linguistic imperialism, morphological types 731
 Linguistic reconstruction, areal linguistics 65
Lingua internacia (Zamenhof) 375
 Li’o 420
see also Flores languages
 Lisboa, Marcos de 158
 Lisu 968–969
see also Lolo-Burmese languages
 Literacy *see* Writing/written language
 Literary Mongol 723
 Literature
 Hebrew 484
 World Englishes 368

- Lithuania, official language 646
 Lithuanian **646–649**
 classification 251–252
 concord 647–648
 dialects 646
 future 648
 grammars (books) 646
 lexicon 648
 morphology 647
 nouns 647
 case forms 647
 gender 647–648
 genitive 647
 as official language 646
 perfect 648
 phonology 646
 consonants 647
 prosody 646
 stress 646
 vowels 647
 verbs 648
 gerunds 648
 participles 648
 tenses 648
 written language 646
 East Prussian tradition 646
 in Grand Duchy 646
 National Standard Language 646
 see also Baltic languages; Balto-Slavic languages; Latvian
 Littoral *Bund* 392–393
 Livonian 1129–1130
 see also Finnic languages
 Llanito *see* Yanito
 Llull, Ramón, artificial languages 76
 Loanword(s)
 Kazakh 590
 Kirghiz 611–612
 Turkic languages 1112
 see also Borrowing
 Local languages, endangerment *see* Language endangerment
 Loewen, Jacob 226
 Logba 631
 see also Togo Mountain languages
 Logic *see* Artificial languages
 Loglan 77
 artificial languages 76
 Logol 613
 see also Kordofanian languages
 Logophoric marking, African linguistic area 5
 Lolo-Burmese languages 968–969
 see also Sino-Tibetan languages
 Lombard *see* Italian
 Lomonosov, Mikhail Vasilyevich 905
 Lomorik 613
 see also Kordofanian languages
 Long-range comparisons **649–656**
 basic vocabulary 650
 borrowing 651
 chance similarities 652
 erroneous morphological analysis 653
 glottochronology 650
 grammatical evidence 651
 hypothesized long-range relationships 649, 649*t*
 lexical comparison 650
 long-range proposals 653
 methods 649
 multilateral comparison 650
 nonlinguistic evidence 652
 nursery forms 652
 onomatopoeia 652
 semantic constraints 652
 short forms and unmatched segments 652
 sound correspondences 650
 sound-meaning isomorphism 652
 spurious forms 653
 Longuda 3
 see also Adamawa-Ubangi languages
 Lontar writing, Balinese 117
 Lottner, C 13
 Louisiana Creole **656–657**
 classification 249*t*
 grammar 657
 Haitian Creole *vs.* 656
 use of 656
 USA 656, 1127
 variation 656
 see also Creoles; Language endangerment; Pidgins
 Lounsbury, Floyd Glenn 808
 Lower Chehalis 749
 see also Salishan languages
 Lozi 1017–1018
 Zambia 1017–1018
 see also Sotho-Tswana languages
 Lü 1039
 see also Tai languages
 Luchuan *see* Ryukyuan
 Lude 1129–1130
 see also Finnic languages
 Ludhianwi Punjabi 886
 see also Punjabi
 Luganda **657–658**
 as agglutinating language 657–658
 concord 658
 future 658
 genetic affiliation 657
 morphology 657
 noun classes 657–658
 nouns 657–658
 orthography 657
 phonology 657
 consonants 657, 657*t*
 syllables 657
 tone 657
 vowels 657, 658*t*
 syntax 658
 use of 657
 speaker numbers 657
 Uganda 657
 verbs 658
 word order 658
 see also Bantu languages; Benue-Congo languages
 Lule Saami 911
 see also Saami
 Luo **658–659**
 Bantu, influence from 659
 classification 253
 downstep 658–659
 grammars (books) 658–659
 noun class prefixes 659
 orthography 658
 SVO 659
 use of 658
 speaker numbers 772–773
 vowel harmony 658–659
 word order 659
 see also Nilo-Saharan languages
 Lushai 968–969
 see also Kuki-Chin languages
 Luvian
 historical aspects 36
 Hittite *vs.* 37
 Luxembourg
 French 427
 German 444
 national language 659–660
 Luxembourghish **659–661**
 adjectives 659–660
 German, related to 659
 history 659
 as national language 659–660
 noun plurals 659–660
 OVS 659–660
 phonology 660
 consonants 660
 sample 660
 SOV 659–660
 syntax 659–660
 third-person pronouns 659–660
 word order 659–660
 see also German; Germanic languages
 Lyngngam 595
 future 596
- M**
 Maasai, Gikuyu, influence on 449–450
 Maban
 classification 773
 geographical distribution 775*f*
 see also Nilo-Saharan languages
 Macalister, R A S 856
 Macbain, Alexander 856
 Macedonia, Republic of
 Albanian 22
 Macedonian 663
 Romani 898
 Romanian 901
 Macedonian **663–665**
 classification 251–252, 974–975
 declensions 976–977
 dialects 664
 genders 664
 imperfective 664–665
 lexicon 664
 morphology 664
 origin/development 663
 Bulgarian 663
 Misirkov, Krste 663–664
 Old Church Slavonic 663
 orthography 664
 Cyrillic alphabet 664
 perfect 664–665
 phonology 664
 resultative 664–665
 syntax 664
 use of 663
 verbs 664–665
 see also Balkan linguistic area; Balto-Slavic languages; Bulgarian; Church Slavonic; Old Church Slavonic; Slavic languages; Slovene
 Macro-Gê 750
 see also Native American languages
 Macro-Jê languages **665–669**
 classification 252–253, 665, 666*t*
 comparative evidence 665
 ergative 668
 inflectional morphology 667
 long-range affiliations 666
 morphology 667
 phonology 667
 vowels 667
 resources 668
 use of 666–667
 geographical distribution 666
 vowel harmony 667
 word order 667–668
 see also Apinajê; Cariban languages; Tucanoan languages; Tupian languages
 Macuna
 accent/tone 1096
 adjectives 1098
 case markers 1097
 consonants 1094
 morphemes 1096
 speaker numbers 1092*t*
 see also Tucanoan languages
 Madagascar
 Austronesian languages 97
 French 674
 Malagasy 674
 Madang languages
 adverbs 671
 case marking 671
 classification 253, 669–670, 1086–1087
 clause chains 671
 complex predicates 671
 development 669
 grammar 671
 nouns 671
 OVS 671
 phonology 670
 consonants 670–671
 syllables 670–671
 vowels 670–671
 resources 669–670
 SOV 671

- structure 670
 suffixes 671
 use of 669
 geographical distribution 669, 670f
 word order 671
 written records 669
see also Papuan languages; Trans New Guinea languages
 Madi 652–653
 Madura 99
 see also Austronesian languages
 Madurese 672–674
 dialects 672
 morphology 673
 phonology 672
 consonants 672, 673t
 vowels 673, 673t
 reduplication 673
 use of 672
 vocabulary 673
 vowel harmony 673
 writing system 673
 see also Austronesian languages; Javanese; Malay
 Maguindanaon
 case-marking 1005, 1005t
 morphosyntax 1005
 phonology 1003
 pronouns 1005, 1005t
 word order 1005
 see also South Philippine languages
 Mahabharata 961
 Mahali 736
 see also Munda languages
 Mahaparitta 832
 Mahavamsa 832
 Mahican 24
 see also Algonquin languages
 Mahl *see* Dhivehi
 Maiduan 750
 see also Penutian languages
 Mailuan languages 1087
 see also Trans New Guinea languages
 Maipure 60
 see also Arawak languages
 Maisica, C P
 Defining a linguistic area 999
 South Asian languages 999, 1000
 Maithili 525–526
 see also Indo-Aryan languages
 Majhi Punjabi 886
 see also Punjabi
 Makah 1157
 vowel epenthesis 1158–1159
 see also Wakashan languages
 Makuráp 1106t
 see also Tupian languages
 Makushi
 gender 184–185
 geographical distribution 185f
 person-marking prefixes 185–186
 see also Cariban languages
 Malagasy 674–677
 classification 250–251
 dialects 674
 genetic relationships 674
 geographical distribution 674, 675f
 morphology 675
 orthography 675, 675t
 phonology 675, 675t
 spirant/stop alternation 675–676
 use of 674
 Madagascar 674
 as official language 674
 verbs 676
 writing systems 674
 Arabic script 674–675
 see also Austronesian languages
 Malawi
 Fanagalo 411
 national language 791
 Nyanja (Chichewa) 791
 Shona 938
 Malay 677–680
 British/Dutch colonization 678
 Indonesian (Bahasa Indonesia) 678
 influence on other languages
 Afrikaans 8, 9, 11
 Tok Pisin 1077
 Malaysian (Bahasa Malayu) 678
 morphology 679
 origin/development 677
 phonology 679
 consonants 679t
 vowels 678t
 use of
 Borneo 678–679
 Indonesia 678
 Malaysia 678
 Papua New Guinea 836
 Singapore 679
 Thailand 679
 see also Austronesian languages;
 Javanese; Madurese;
 Malayo-Polynesian languages;
 Riau Indonesian
 Malayal, morphological causatives 997
 Malayalam 680–684
 classification 251
 consonants 298
 dialects 681
 etymology/variant names 680
 future 683
 genetic affiliation 682
 grammar 681
 influence from other languages 680–681
 interrogative sentences 683
 literature development 680
 morphology 682
 morphophonemics 682
 sandhi rules 682
 nominative case 682
 nouns 682
 accusative 301–302
 noun phrases 683
 numerals 682
 personal suffixes 305
 phonology 682
 single velar stops 682
 voiceless stops 682
 vowels 682
 pronouns 302–303, 303t
 script 297, 681
 Vatteluttu 681
 see also Tamil script
 sentences 683
 SOV 683
 stops 683
 syntax 683
 three-way tense distinctions 683
 use of 680
 verbs 682–683
 finite verbs 304
 vowel-ending stems 683
 workers in 681
 see also Dravidian languages; Kannada; Tamil; Telugu
 Malayo-Polynesian languages 684–688
 classification 250–251, 685f, 686f
 definition 684
 development 684
 ‘politeness shift,’ 684
 prefixes 684–685
 pronouns 684
 settlement 687
 history 685
 archaeological evidence 685
 integrity 684
 Mon-Khmer languages, influence from 687
 resources 687
 structure 687
 subgrouping 685
 use of 684
 workers in 684
 see also Austronesian languages; Balinese; Bikol; Flores languages; Formosan languages; Kapampangan; Malay; Malukan languages; Maori; North Philippine languages; Papuan languages; Proto-Austronesian; Samar-Leyte; South Philippine languages
 Malaysia
 Asian languages 94–95
 Austronesian languages 97, 99
 English 678
 Hindi 495
 Malay 678
 national language 678
 official language 678
 Malaysian (Bahasa Malayu) 678
 Maldives Islands
 Dhivehi 285
 Indo-Aryan languages 522
 official language 285
 Maldivian
 classification 522
 writing systems 524
 Tana (Thaana) 524
 see also Indo-Aryan languages
 Malecite-Passamaquoddy
 classification 24
 speaker numbers 26
 see also Algonquin languages
 Mali
 Berber 152
 Dogon 771
 Gur languages 472
 Mande languages 769–770
 Songai languages 990–991
 Malta
 English 688
 Italian 545
 Maltese 688
 official languages 688
 Maltese 688–689
 see also Afroasiatic languages
 Malukan languages 689–691
 alienable-inalienable contrasts 690
 classification 250–251
 as endangered languages 690
 SVO 690
 use of 689, 689f
 speaker numbers 690
 word order 690
 see also Austronesian languages;
 Central Malayo-Polynesian (CMP);
 Language endangerment;
 Malayo-Polynesian languages; Papuan languages
 Malwi Punjabi 886
 see also Punjabi
 Mam 705–706
 dialects 705–706
 positionals 707t
 speaker numbers 706t
 transitive verbs 708
 see also Mayan languages
 Mambila 691–693
 classification 253, 691
 dialects 1213
 language vitality 692
 morphology 692
 phonology 691
 consonants 691
 tone 692
 vowels 692
 plurals 1213
 suffixes 1213
 SVO 692
 syntax 692
 use of 1213
 Cameroons 691
 Nigeria 691, 1213
 speaker numbers 1213
 word order 1213
 see also Bantu languages; Benue-Congo languages
 Mamean 705–706
 see also Mayan languages
 Mampruli 472
 see also Oti-Volta languages

- Manambu 693–694
 classification 253
 as agglutinating language 693
 as endangered language 694
 as synthetic language 693
 clause-chaining 693
 dialects 693
 genders 693
 morphology 693
 nouns 693
 numbers 693
 personal name ownership 693–694
 phonology 693
 consonants 693
 vowels 693
 switch-reference 693
 use of 693
 verbs 693
see also Ndu languages; Tok Pisin
- Manchu, Korean 614
- Manda 300*t*
see also Dravidian languages
- Mandarin Chinese 214
 classification 969
 classifiers 1013
come to have verb 1015
 indeterminateness 1011
 lexical tones 223
 use of 969
 speaker numbers 214*t*
 word order 1014*t*
see also Chinese; Sinitic languages; Sino-Tibetan languages
- Mande languages 769
 classification 768–769
 current 696
 Delafosse, Maurice 696
 earlier 696
 Greenberg, Joseph H 696–697
 Mukašovsky, Jan 696–697
 Steinthal, Heymann 696
 vocabulary 697
 Westerman 696–697
 Williamson 696–697
 diminutives 698
 Hausa, influence on 477
 morphosyntax 698
 noun classes 698
 phonology 697
 consonants 697
 reconstructed history 694
 linguistic evidence 695–696
 resources 698
 SOV 697
 tone 697
 use of 769–770
 speaker numbers 694
 variants 694*t*
 verb suffixes 698
 word order 698
 workers in 696
see also Niger-Congo languages
- Mande Studies Association (MANSA) 698–699
- Mandingo, Krio, influence on 620
- Mangala *see* Jiwari
- Manggarai
 classification 420
 voice alteration 420
see also Flores languages
- Mang languages 729
see also Viet-Muong languages
- Manichaeism 531–532
- Manimekalai* 1047–1048
- Manjaku-Papel 770
see also Atlantic Congo languages
- Mano 697–698
see also Mande languages
- Manobo languages
 classification 1001–1002, 1002*t*
 consonants 1003
 contrast neutralization 1003
see also South Philippine languages
- Mansi
 classification 1129–1130
 diphthongs 1130
 object marking 1132
 vowel harmony 1130
see also Uralic languages
- Mansim
 classification 1176
see also East Bird's Head (EBH) languages
- Manubaran languages
 classification 1087
see also Trans New Guinea languages
- Manx
 classification 200, 251–252
 development 454
 sample 454
see also Goidelic Celtic; Goidelic languages
- Manyika 1017
see also Shona languages
- Mao 805
see also Omotic languages
- Maori 699–701
 classification 250–251
 dialects 700
 lexicon 700–701
 long-range comparisons 651
 Niuean, influence on 776
 as official language, New Zealand 700
 phonemes 700
 revival of 700
 syntax 700
 VSO 700
see also Malayo-Polynesian languages; Oceanic languages
- Mapoyo 185*f*
see also Cariban languages
- Mapuche 41
 Chile 41
 Patagonia 41
 verbs 41
see also Andean languages
- Mapudungan languages 701–703
 affiliations 701
 classification 252–253, 701
 influence from other languages 702
 iterative 702
 lexicon 702
 noun morphology 702
 noun phrases 702
 phonology 701–702
 consonants 701–702
 fricatives 701–702
 vowels 701–702
 pronouns 702
 resultative 702
 use of 701
 verb morphology 702
- Maraya 112–113
see also Azerbaijani
- Marand 112–113
see also Azerbaijani
- Maranungku 89
see also Australian languages
- Marathi 703–705
 classification 251–252, 522
 compound verbs 996
 correlative 704
 ergative 524
 history 703
 morphology 704
 notion of subject 704
 nouns 704
 passivization 704
 phonology 525–526, 703
 alphabet 703
 consonants 703, 704*t*
 vowels 703, 703*t*
 script 703
 SOV 704
 subordination 704
 suprasegmentals 703
 accent 703
 nasal vowels 703
 syntax 704
 use of 703
 number of speakers 523
 word order 704
 writing systems, Devanagari 524
see also Dardic languages; Indo-Aryan languages
- Margany
 case marking 87, 87*t*
 morphology/syntax 87
 verbs 87–88
see also Australian languages
- Margay 90
see also Australian languages
- Mari languages
 classification 1129–1130
 object marking 1132
 vowel harmony 1130
 word order 1132
see also Uralic languages
- Marking, in isolating language 222
- Maronites 1033
- Marriage aspects, Tucanoan languages 1092
- Masatekan languages
 classification 819–821
 syllable onsets 821–822
see also Oto-Mangean languages
- Masateko 819
 time depth 819
see also Oto-Mangean languages
- Masawa 819–821
see also Oto-Mangean languages
- Mason, J Alden 1074
- Masorettes, Hebrew 483
- Matagalpa
 classification 711
 use of 711
see also Misumalpan languages
- Matbat 1177
see also West Papuan languages
- Mathews, R H 438
- Matlatzinka 751
 classification 819–821
see also Oto-Mangean languages
- Mator 1129–1130
see also Samoyed languages
- Matteson, Ester 60
- Mauritania
 Arabic 42
 Fulfulde 430
 Mande 769–770
 official language 42
 Wolof 1184
- Mauritius, Hindi 495
- Mawé
 classification 1106*t*
 ideophones 1106–1107
 positional demonstratives 1106
see also Tupian languages
- Maxakali
 classification 665, 666, 666*t*
 ergative 668
 morphology 667
 as ergative language 668
 use of 667
 geographical distribution 666–667
see also Macro-Jê languages
- Maxi (Maxi-gbe) 631–632
see also Gbe languages
- May 728–729
see also Chut languages
- Ma'ya 1177
see also West Papuan languages
- Mayali *see* Gunwinygu (Gunwigu)
- Mayan languages 751
 affiliations, Mapudungan languages 701
 classification 705–706
 classifiers 708, 708*t*, 709*t*
 directional particles 708
 ergative 707
 geographical distribution 705
 grammar 707
 influence on other languages 708–709
 influences from other languages 708–709
 noun classifiers 708
 numeral classifiers 708
 positionals 707

- transitive verbs 708
 use of 705–706
 Guatemala 705–706, 706*t*
 Mexico 707*t*
 as official language 705–706
 speaker numbers 706–707
 verbs 707
 vocabulary 708
see also Achi'; Akateko; Awakateko; Native American languages
 Mayan-Mixe-Zoquean languages 748
see also Native American languages
 Maybrat 1176
see also West Papuan languages
 Mazatecan 751
 Mazatlán *see* Mixe-Zoquean languages
 Mazdaism *see* Zoroastrianism
 Mbatto 631
 Mbum languages 3
see also Adamawa-Ubangi languages
 Meadow (Eastern) Mari 1129–1130
see also Mari languages
 Media, language endangerment role 324
 Medio-passive, Old Irish 453
 Mediterranean language area 392–393
 Medlpa
 classification 1086–1087
 speaker numbers 836, 1085
see also Chimbu-Wahgi languages; Papuan languages; Trans New Guinea languages
 Meghalaya 595
 Megleno-Romanian dialect 901
 Meherrin 542
see also Iroquoian languages
 Méhif (Michif) *see* Michif
 Meillet, Antoine Paul Jules, Balto-Slavic languages 135–136
 Meinhof, Carl Friedrich Michael, Hamitic theory 13
 Meinhof's law, Bantu languages 139
 Mekéns
 classification 1106*t*
 positional demonstratives 1106
see also Tupian languages
 Mek languages 1085–1086
see also Trans New Guinea languages
 Mel languages 770
see also Atlantic Congo languages
 Mendi (Angal Enen)
 classification 1086–1087
 Krio, influence on 620
see also Engan languages
 Menomini
 classification 25
 speaker numbers 26
see also Algonquin languages
 Mercian (midlands), Old English dialect 356
 Meryam Mer 79
 Mesoamerica 63
 Mesolect 868
 Métchif *see* Michif
 Mexicano *see* Nahuatl
 Mexico
 Mayan languages 707*t*
 Tohono O'odham 1074
 Totonacan languages 1080
 Meyah
 classification 1176
 nominal complex, numbers 1177
 tone 1177
see also East Bird's Head (EBH) languages; West Papuan languages
 Meyaña 112–113
see also Azerbaijanian
 Michif 261, 709–711
 agreement 710
 influence from other languages
 Cree 710
 English 710
 French 710
 origin/development 28, 709–710
 orthography 710
 phonology 710
 use of 709
 verbs 710
 word order 710
see also Algonquin languages; Ritwan languages
 Micmac
 classification 24
 speaker numbers 26
see also Algonquin languages
 'Micro-Altaic,' 30
 Middle America, Native American languages 748
 Middle Arabic 932
 Middle Aramaic 57–58, 934
 Middle Egyptian 39
 Middle English 351–356
 declensions 353
 definition 351
 determiners 353
 development to Early Modern English *see* English, Early Modern
 dictionaries 355
 example 354
 external history 351
 grammar 353
 graphology 352
 alphabet 352
 inflection 353
 internal history 352
 lexicon 354
 French loans 352, 354
 Latin loans 352, 354
 Norse loans 354
 modern works 355
 noun phrase 353
 origin/development 351
 phonology 353
 consonants 353
 diphthongs 353
 long vowels 339
 unstressed vowels 353
 vowels 353
 SOV 354
 verb phrase 354
 word order 354
see also English, Early Modern; English, Later Modern; English, Modern; French; Germanic languages; Icelandic; Latin; Norse, Old; Old English
 Middle English dictionary 355
 Middle Hebrew 934
 Middle Persian *see* Pahlavi (Middle Persian)
 Middle Wahgi 1086–1087
see also Chimbu-Wahgi languages
 Midrashim, Jewish Palestinian Aramaic 58
 Mienic 503
see also Hmong-Mien (Miao-Yao) languages
 Mikasuki 738–739
 agreement type 740–741
 vowel length 740
 Milindapañha 832
 Min
 classification 969
 geographical distribution 969
see also Sinitic languages; Sino-Tibetan languages
 Minaean 931
see also Semitic languages
 Minangkabau 99
see also Austronesian languages
 Min languages 219
 classification 214
 speaker numbers 214*t*
see also Chinese; Sinitic languages
 Minority language endangerment *see* Language endangerment
 Miri 613
see also Kadugli languages
 Misanla Totonac 1080
 applicative affixes 1083–1084
 body part prefixes 1083
 nouns 1083
 object agreement 1084
 phonology
 consonants 1082
 vowels 1082
 use of 1080–1081
 word order 1084
see also Totonacan languages
 Misher Tatar 1052
 Mising (Miri) 968–969
see also Tani languages
 Misirkov, Krste 663–664
 Miskito
 classification 711
 use of 711
see also Misumalpan languages
 Misteko 819
 classification 819–821
 time depth 819
see also Oto-Mangean languages
 Misumalpan languages 711–712
 classification 252
 use of 711
 classification 711
see also Chibchan languages
 Miwok-Costanoan 750
see also Penutian languages
 Mixed languages *see* Creoles; Pidgins
 Mixe-Zoquean languages 751
 aspects 714
 auxiliary constructions 715
 classification 712
 as agglutinative languages 714
 cliticization 714
 dependent verb forms 715
 derivational nouns 714
 future 712
 inflectional classes 714
 inversive person marking 714–715
 long-range comparisons 653
 morphology 714
 nouns 714
 origin/development 712
 persons 714
 phonology 712
 consonants 712
 glottal stop metathesis 713
 syllable cods 713
 unstressed vowel loss 713
 positional references 715
 syntax 715
 transitive verbs 714–715
 verbs 714
 VSO 715
 word order 715
see also Native American languages
 Mixtec 751
see also Mixtecan languages
 Mixtecan
 endocentric noun classes 823
 noun classes 823
 syllable onsets 821
 Mixtecan languages 751
see also Cuiclatec
 Mnong 726
see also Bahnaric languages
 Moabite 934
 Phoenician *vs.* 854
see also Semitic languages
 Mobilian Jargon (Mobilier) 716–718, 739
 case marking 716–717
 classification 249*t*
 definition 857–858
 development 716–717
 influence from other languages 716–717
 origin 717
 use of, social context 717
 word order 716–717
see also Algonquin languages; Creek; Creoles; Muskogean languages; Pidgins; Ritwan languages; Siouan languages
 Mochica 41
 Mocho' 705–706
 speaker numbers 707*t*
see also Mayan languages
 Modality
 Evenki 407
 Nuristani languages 787
 sign language 946, 946*f*
 Wolaitta 1182

- Modern Persian *see* Persian, Modern
 Modern Southern Arabian 931
see also Semitic languages
 Modern Standard Arabic (MSA) 43
 influence from other languages 43
see also Arabic; Aramaic; Syriac
 Modern Standard Chinese *see* Putonghua
 Modern Tiwi 1065–1066, 1067
 Moghol 722
see also Mongol languages
 Mohawk
 noun incorporation 544
 subject/object categories 544
 tone 543
 use of 543
 verbs 543–544
see also Iroquoian languages
 Molale 750
see also Penutian languages
 Moldova, Gagauz 1112
 Molo 772–773
see also Nilo-Saharan languages
 Mon 718–721
 causative verbs 725
 classification 250
 as tonal language 719
 decline of 718
 dialects 718
 dictionaries 718–719
 example 720
 Internet 719
 phonology 719
 consonantal clusters 719–720, 720*t*
 consonants 719, 719*t*
 minimal pairs 719, 719*t*
 syllable rhythms 720
 vowels 719, 720*t*
 scripts 719
 SVO 720
 use of
 Burma/Myanmar 718, 727
 speaker numbers 718
 Thailand 718, 727
 written texts
 oldest 718, 719
 spoken *vs.* 718
see also Austroasiatic languages; Monic languages; Mon-Khmer languages
 Monaco, Italian 545
 Mondé
 adjectives 1106
 classification 1106*t*
 ideophones 1106–1107
 tone system 1106
see also Tupian languages
 Mongolia
 Evenki 405
 Kazakh 588
see also Altaic languages; Areal linguistics; Classification (of languages); Kazakh; Language endangerment; Tungusic languages; Turkic languages; Uralic languages
 Mongolian (Khalkha) *see* Khalkha
 Mongolian, Inner 969
see also Jin languages
 Mongol languages 721–724
 classification 250, 722, 722*t*
 definition 721
 demography 723
 distribution 721
 genetic status 722
 literary use 723
 political status 723
 research history 723
 time depth 722
 Turkic-Tungusic relationship 30
 types 722
 use of 723
see also Altaic languages
 Monic languages 724
 use of 727
see also Mon-Khmer languages
 Mon-Khmer languages 95, 724–730
 causative verbs 725
 classification 250, 724
 Malayo-Polynesian languages, influence on 687
 morphology 725
 phonology 725
 serial verb constructions 725
 subgroupings 1010*t*
 syntax 725
 use of 725
 verbs 725
see also Austric hypothesis; Austroasiatic languages; Khasi languages; Khmer (Cambodian); Mon; Southeast Asian languages; Vietnamese; Wa
 Monomorphemic signs, sign language, morphology 949, 949*f*
 Monomorphemic words, in isolating language *see* Chinese
 Montagnais
 classification 24–25
 phonology, stress 26
see also Algonquin languages
 Montenegro, Republic of 22
 Monumbo 1078
see also Torricelli languages
 Mood
 Bislama 162
 Indo-Iranian languages 534
 Nenets (Yurak) 762–763
 Spanish 1021
 Telugu 1057
 Wolaitta 1183, 1183*t*
 Xhosa *see* Xhosa
 Mòoré 472
see also Oti-Volta languages
 Mopan 705–706
 speaker numbers 706*t*
see also Mayan languages
 Mordvin languages
 classification 1129–1130
 phonology
 vocalism 1130
 vowel harmony 1130
 word stress 1131
 vowel harmony 1130
see also Uralic languages
 Moreno, W W 378–379
 Moribund languages
 definition 319–320
 reversal/revitalization strategies 326
see also Language endangerment
 Moro 613
see also Kordofanian languages
 Morocco, Berber 152
 Morphemes
 in isolating language 222
 morph ratio 731
 Morphological causatives, South Asian languages 997
 Morphological types 730–735
 clustering features 731
 agglutinating type 732
 comparisons 733, 733*t*
 fusional type 732
 isolating type 732
 contemporary 733
 morphology-syntax interface 734
 partial 733
 universals 734
 word order importance 733–734
 criticism of 730–731
 Greenberg 731
 linguistic imperialism 731
 Sapir 731
 morpheme-morph ratio 731
 morph word internal modification 731
 origin/development 730
 Greenberg, Joseph 730
 von Humboldt, Wilhelm 730
 Sapir, Edward 730
 Schlegel, August Wilhelm 730
 Schlegel, Friedrich 730
 postposition 733–734
 ratio of morph to word forms 731
 types 730
 agglutinating 731
 classical 731
 continuum 731
 fusional 731
 ideal 731
 isolating 731
 workers in
 Greenberg, Joseph 730
 von Humboldt, Wilhelm 730
 Sapir, Edward 730
 Schlegel, August Wilhelm 730
 Schlegel, Friedrich 730
 Morphological universals 734
 Morphology
 gender assignment 758
 postbases 202, 203*t*
 syntax interface 734
see also specific languages
 Morrobalama 735–736
 as endangered language 735
 ergative 735
 phonology 735
 changes 735
 split-ergative system 735
 suffixes 735
 use of, Australia 735
 word order 735
see also Arrernte; Australian languages; Creoles; Pidgins
 Moru 652–653
 Mosetén 41
see also Andean languages
 Mosina dialect *see* Vurës
 ‘Mother-in-law’ languages, Australian languages 91
 Motu *see* Hiri Motu
 Motuna (Siwai) 841
see also South Bougainville languages
 Mouthing *see* Sign language(s)
 Movima 41
see also Andean languages
 Mozambique
 Fanagalo 411
 Nyanja 791
 Portuguese 883
 Shona 938
 Shona languages 1017
 Southern Bantu languages 1017
 Swahili 1026
 Tshwa 1018
 Tsonga 1018
 Mpi 968–969
see also Lolo-Burmese languages
 Mpur
 classification 1176
 tone 1177
see also West Papuan languages
 Mudari *see* Bhumij (Mudari)
 Mudo 613
 Muisca *see* Chibchan languages
 Mukašovský, Jan, Mande language
 classification 696–697
 Multilingualism
 language shift 326
see also Bilingualism
 Mumuye languages 3
see also Adamawa-Ubangi languages
 Munda languages 94, 736–738
 classification 250
 contacts 737
 morphology 737
 Northern group 736
 morphology 737
 phonology 737
 possessives 737
 Southern group 736
 morphology 737
 SOV 736–737
 SVO 737
 use of 736
 verb construction 737

- VSO 737
see also Agariya; Asuri; Austroasiatic languages; Santali
- Mundari 736
see also Munda languages
- Mundurukú
 classification 1106*t*
 ideophones 1106–1107
 noun classification 1108
 positional demonstratives 1106
 tone system 1106
see also Tupian languages
- Munsee 24
see also Algonquin languages
- Munster, Irish, development of 454
- Muong languages 728–729
 speaker numbers 728–729
see also Viet-Muong languages
- Muskogean languages 749
 agreement type 740–741
 classification 739
 morphology 740
 noun phrases 740
 phonology 739
 consonants 739, 739*t*
 pitch-accent 739–740
 suffixes 740
 vowel length 740
 vowels 739
 postpositions 741
 types 738–739
 use of 738
 verbs 740, 741
 word order 741
see also Alabama-Koasati; Atakapa; Timucua; Tunica; Yuchi; Yukian
- Muskogee 754
- Mutation
 Breton 167
 Finnish (Suomi) 418
 Nivkh 778
 Old English 358–359
 Slovak 979
 Welsh 1170
- Mutual intelligibility, North American native language variation 754
- Muya 968–969
see also Qiangic languages
- Muzo *see* Cariban languages
- Myanmar *see* Burma/Myanmar
- Mycenaean Greek 461
- N**
- Na-Dene languages 748
 and Amerind 655
 classification 743
 classifiers 743–744
 distribution 743
 historical aspects 747–748
 types 743, 744*t*
see also Athabaskan–Eyak–Tlingit (AEC); Native American languages
- Nagamese 78
- Naga Pidgin 78
- Nagara Apabhramsa 468
- Nagovisi 841
see also South Bougainville languages
- Nahali (Nihali) 94
see also Austroasiatic languages
- Nahua *see* Nahuatl
- Nahuatl 745
 classification 252
 dialects 745
 dictionaries 745
 grammars (books) 745
 history 745
 Mayan languages, influence on 708–709
 nouns 745
 phonology 745
 consonants 745
 vowels 745
- use of 745
 word order 745
 workers in 745
see also Central Siberian Yupik; Polysynthetic languages; Uto-Aztecan languages
- Nakh-Daghestanian
 morphology 194
 vowels 193, 193*t*
 word order 195
see also Caucasian languages
- Nama *see* Khoekhoe
- Naman language 101
see also Austronesian languages
- Names
 Etruscan 388
 Fulfulde taboos 433
- Namibia
 Afrikaans 7
 Fanagalo 411
 Khoekhoe 601, 602
- Namuyi 968–969
see also Qiangic languages
- Nance, Robert Morton 258
- Narragansett 24
see also Algonquin languages
- Nasal assimilation, Tucanoan languages 1095
- Nasal spreading, Tucanoan languages 1092
- Nasioi 841
see also South Bougainville languages
- Naskapi 24–25
see also Algonquin languages
- Nasu 968–969
see also Lolo-Burmese languages
- Native American languages 746–753
 areal analysis 746
 diminutives 757
 discourse morphology 746
 geographic distributions 746
 historical aspects 747
 Brinton, Daniel G 747
 Henshaw, H W 747
 re-classification 748
 linguistic features 746
 linguistic stock identification 747
 vocabulary inspection 747
 middle America 748
 north American isolates resisting affiliation 751
 polysynthesis 746
 sentence morphology 746
 social distributions 746
 South America 748
 Andean area 752
 lowlands 752
 southern ‘cone’ area 752
 southern Texas/Mexico isolates 751
see also Aboriginal languages; Algonquin languages; Arana; Araucanian; Arawakan; Aymara; Coahuilteco; Karankawa; North American native languages; Oto-Mangean languages; Panoan languages; Puquina; Quechua languages; Salishan languages; Siouan languages; Solano
- Nativization, World Englishes 365
- Nauatl *see* Nahuatl
- Navajo 761
 classification 252
 language shift 319, 323
 postpositions 761
 SOV 761
 speaker numbers 323
 syllables 761
 verbs 761
 vowels 761
 word order 761
see also Na-Dene languages
- Nawat *see* Nahuatl
- Nawuri 631
see also Guang languages
- Nchumuru 631
see also Guang languages
- Ndau 1017
see also Shona languages
- Ndebele 1017
 South Africa 1187
 Zulu *vs.* 1215
see also Nguni languages
- Nding 613
see also Kordofanian languages
- Ndu 1078
- Ndu languages 253
see also Papuan languages
- Nederlandsche Taal-en Letterkundig Congressen* 308
- Nedungadi, Kovunni 681
- Negation
 future tense 129*t*
 sign language affixation 949, 950*f*
- Negative pronouns 393–394
- Nembe–Akaha 517
- Nen 137
see also Bantu languages
- Nenets (Yurak) 761–764
 classification 1129–1130
 as agglutinating language 762
 as endangered language 763
 as synthetic language 762
 conjugation 762–763
 Forest Nenets 761–762
 future 762–763
 imperfective 762–763
 literary tradition 763
 moods 762–763
 nouns 762
 phonology
 consonants 762
 glottal stops 762
 vowels 762
 word stress 1131
 possessives 762
 postpositions 762
 pronouns 762, 763
 resources 763
 Russian bilingualism 761–762
 SOV 763
 tenses 762–763
 Tundra Nenets 761–762, 762–763
 use of 761–762
 speaker numbers 761–762
 verbs 762–763
 nonfinite verbs 763
 word order 763
see also Samoyed languages
- Neo-Aramaic *see* Aramaic, Modern (Neo-Aramaic)
- Neogrammarianism, Indo-European languages 528–529
- Nepal
 Hindi 495
 Indo-Aryan languages 522
 Munda languages 736
 Nepali 764
 Tibetan 1060–1061
- Nepali 764–765
 Bengali *vs.* 764
 classification 251–252
 classifiers 764
 dialects 764
 Hindi *vs.* 764
 literary purposes 764
 number of speakers 523
 tonal system 525–526
 use of 764
 writing systems, Devanagari 524
see also Dardic languages; Indo-Aryan languages
- Nera, use of, distribution 775*f*
- The Netherlands, Dutch 307
- New Caledonia, Javanese 560
- ‘New language,’ North American native language variation 755
- New Persian *see* Persian, Modern
- New Tiwi *see* Tiwi

- New Zealand
 Fijian 412
 Maori 700
 Ngad'a 420
see also Flores languages
 Nganasan (Tavgy)
 classification 1129–1130
 phonology
 diphthongs 1130
 vowel harmony 1130
 word stress 1131
see also Samoyed languages
 Ngan'gi 765–768
 classification 765*t*
 classifiers 766
 morphology 766
 noun classes 766–767
 phonology 767
 consonants 767*t*
 vowels 767*t*
 as polysynthetic language 765–766
 pronominal indexing 766
 pronouns 766, 767*t*
 use of 765
 speaker numbers 765
 verbal classification 766
see also Australian languages;
 Central Siberian Yupik; Polysynthetic
 languages
 Ngbaka-Mba languages 3
see also Adamawa-Ubangi languages
 Ngbandi languages 3
see also Adamawa-Ubangi languages
 Ngile 613
see also Kordofanian languages
 Nguni languages 1017
 Fanagalo, influence on 412
see also Bantu languages, Southern
 Nguôn 728–729
see also Muong languages
 Nha Hôn 726
see also Bahnaric languages
 Niaibo 624
see also Grebo languages
 Nicaragua
 Arawak languages 59
 Misumalpan languages 711
 Sumu (Sumo Tawahka) 711
 Nichols, Johanna 529
 Nicobarese 94
 Austric hypothesis 92
 influence from other languages 93
 noun phrases 92–93
 text materials 92–93
 word order 92–93
see also Austroasiatic languages
 Nicolaisen, W F H 856
 Nida, Eugene Albert, Inupiaq
 writing 535–536
 Niger
 Berber 152
 Kanuri 578
 Mande 769–770
 Songai languages 990–991
 Niger-Congo languages 768–772
 accepted classification 769*f*
 classification 253
 early classification 768
 Bleek 768
 Greenberg, Joseph H 769*f*
 Koelle, Sigismund W 768
 Westermann 768
 Krio, influence on 617
 Nilo-Saharan languages *vs.* 773–774
 use of 768
see also Adamawa-Ubangi languages;
 Akan; Atlantic Congo languages;
 Bantu languages; Benue-Congo
 languages; Dogon; Efik; Ewe; Gur
 (Voltaic) languages; Kordofanian
 languages; Kru languages; Kwa languages;
 Mande languages; Nilo-Saharan
 languages; Nyanja; Swahili; Wolof;
 Xhosa; Yoruba
 Nigeria
 Adamawa-Ubangi languages 771
 Arabic 42
 Benue-Congo 771
 Efik 314
 Gur languages 770
 Hausa 477, 1209
 Igbo 1209
 Ijò 517
 Kanuri 578
 Kwa 771
 Mambila 691
 Mande languages 769–770
 Nilo-Saharan languages 774
 Yoruba 1207
 Nikayas 832
 Nilo-Saharan languages 772–776
 Afroasiatic languages *vs.* 773–774
 case suffixes 774–775
 classification 253
 consonants 774
 converbs 774
 downstep 774
 grammars (books) 773
 Niger-Congo languages *vs.* 773–774
 noun referring 774
 orthography 773
 subgroups 773*t*
 tone 774
 use of 772
 Chad 774
 distribution 775*f*
 Eritrea 774
 Ethiopia 774
 Nigeria 774
 speaker numbers 772–773
 Sudan 774
 vowel harmony 773–774
 workers in, Greenberg, Joseph H 4, 772
see also Afroasiatic languages; Aka;
 Dinka; Kanuri; Khoesaa languages;
 Luo; Niger-Congo languages; Songhay
 languages
 Nilotic languages
 classification 773
 use of 775*f*
 vowel harmony 773–774
see also Nilo-Saharan languages
 Nimbari 3
see also Adamawa-Ubangi languages
 Nimuendaju, Curt, Macro-jê language
 classification 665–666
 Nipmuck 24
see also Algonquin languages
 Nisu 968–969
see also Lolo-Burmese languages
 Niuean 776–777
 ergative 776–777
 influence from other languages 776
 lexicon 776–777
 phonology 776–777
 as split-ergative language 776
 VSO 776
 word order 776
see also Tongan
 Nivkh 777–779
 case forms 778
 classification 249
 counting forms 778
 future 778
 morphology 778
 mutation 778
 subordinate clauses 778
 use of 777–778
 velar nasal 778
see also Turkic languages
 Nkonya 631
see also Guang languages
 Nkoro 517
 Noble, G Kingsley, Arawak
 languages 60
 Nocte 968–969
see also Konyak languages
 Noghay 589
 Nominals 398–399
 affixes 290
 in agglutinating languages 416*t*
 Old Irish 453
 sign language 950*f*
 Nominative case
 in agglutinating languages 416*t*
 Chuvash 245
 Oromo 810–811
 Nonobligatory categoricals, Southeast Asian
 languages 1011*t*
 Non-Pama-Nyungan languages 89
see also Australian languages
 Nonpolysynthetic languages, productive
 affixation 203–204
 Nootka *see* Nuuchahnulth; Nuuchahnulth
 (Nootka)
 Nootkan branch, Wakashan
 languages 749–750
 Norse, Old
 influence on other languages
 French 248
 Goidelic languages 453–454
 Middle English 354
 Old English *vs.* 356–357
 pro-drop 781
see also Danish; Germanic languages;
 Icelandic; Indo-European languages;
 Middle English; Norwegian;
 Swedish
 North Alaskan Inupiaq (NAI) 535
see also Inupiaq
 North American native languages
 augmentative 757
 gender indicators 757
 meanings 758
 morphology 758
 phonology 757
 vocabulary 758, 758*t*
 language density 319
 lexicostatistics 248–249
 Northwest coast, as linguistic
 area 63
 use of, USA 1127
 variation 753–761
 abnormalities 756
 assessment 754
 baby talk 757
 caretaker language 757
 folk characters 756
 folktales 759
 grammar 755, 756*t*
 intergenerational 755
 lexicon 755
 mutual intelligibility 754
 new language 755
 personal indicators 756
 phonology 755, 755*t*
 regional dialects 754
 research history 753–754
 style 758
see also Algonquin languages; Hopi; Lakota;
 Language endangerment; Muskogean
 languages; Pomoan languages; Ritwan
 languages
see also Canada; USA
 North Arabian *see* Arabic, North
 North Bougainville languages 841
see also Papuan languages
 North Caucasian languages 251
see also Abkhaz; Caucasian languages
 Northeastern Mandarin
 classification 214
 speaker numbers 214*t*
see also Mandarin Chinese
 Northern Brahmi script *see* Brahmi script
 Northern Qiang
 classification 968–969
 morphology 970
 syllable structure 969–970
see also Qiangic languages
 Northern Sotho 1017
 South Africa 1017–1018
see also Sepedi

- Northern Totonac 1080
 speaker numbers 1081
 vowels 1082
see also Totonacan languages
- North Germanic languages *see* Germanic languages, North
- North Gurage languages 382–383
see also Ethiopian Semitic languages
- North Halmahera languages
 classification 1176
 word order 1176
see also West Papuan languages
- North Philippine languages 783–785
 development 783
 ergative 784
 intransitive constructions 784
 noun phrases 783–784
 phonology 784
 affricate prevocalic stops 784
 fricatives 784
 vowels 784
 resources 784
 syntax 783–784
 use of 783
 speaker numbers 783
see also Austronesian languages; Bikol; Ilocano; Kapampangan; Malayo-Polynesian languages; South Philippine languages; Tagalog
- Northumbrian (Northern) dialect 356
- North West Greenlandic 1172
see also West Greenlandic
- Northwest Semitic languages 932
 unidentifiable 934
see also Semitic languages
- Norway
 Old Icelandic 779
 Saami 911
- Norwegian 785–786
 adverbs 785–786
 auxiliary verbs 785
 classification 251–252
 interrogative clauses 786
 nouns 785
 passives 786
 phonology 785
 Russenorsk, influence on 904
 sociohistorical setting 785
 Aasen, Ivar 785
 subject requirements 786
 SVO 785
 syntax 785
 verbs 785
 second language 785
 word order 785
 workers in, Aasen, Ivar Andreas 785
 written standards 785
see also Danish; Germanic languages; Icelandic; Norse, Old; Scandinavian languages; Swedish
- Nostratic theory 786–787
 Afroasiatic languages 786
 Altaic languages 249, 786
 Berber languages 249
 culture 786
 Dravidian languages 249
 Indo-European languages 249, 530, 786
 Japanese 249
 Kartvelian languages 249, 786
 Korean 249
 lexical terms 786
 long-range comparison 649, 652
 methodological problems 654–655
 parent languages 786–787
 Semitic languages 249
 Uralic languages 249
 workers in 249
- Nottoway 542
see also Iroquoian languages
- Noun(s)
 Anatolian languages 37
 Hurrian 516
 in inflecting language 50
 in polysynthetic languages 203
- Tanoan languages 1049–1050
 Thai 1059
- Noun classes
 Adamawa-Ubangi languages 3
 Australian languages 89
 Bantu languages 140
 Bantu languages, Southern 1018
 Benue-Congo languages 151
 Burushaski 176
 Creoles 862
 Fulfulde 430
 Gikuyu (Kikuyu) 450
 Ijò 518
 Ket 593
 Khasi languages 596
 Kordofanian languages 613
 Kru languages 624
 Lak 636
 Luganda 657–658
 Mande languages 698
 Mixtecan 823
 Ngan'gi 766–767
 Nyanja 796–796
 Oaxacan languages 823
 Oto-Mangean languages 823
 Papuan languages 842
 Sepik-Ramu languages 842
 Shona languages 938
 Swahili 1027
 Torricelli languages 842
 Tswana (Setswana) 1018–1019
 Yoruba 1208
 Zapotecan languages 823
 Zulu 1017
- Novgorodov, S A, Yakut 1200
- Nubian
 classification 773
 use of 775f
 vowel harmony 773–774
see also Nilo-Saharan languages
- Nukha 112
see also Azerbaijani
- Numbers, sign language morphology 949–950, 950f
- Numic 1139t
see also Uto-Aztecan languages
- Numidian script *see* Berber
- Nung 1039
see also Tai languages
- Nupoid 151
see also Benue-Congo languages
- Nuristani languages 787–788
 area spoken 787
 as ergative language 788
 modal forms 787
 numbers speaking 787
 phonetics 787
 spatial orientation 788
 temporal forms 787
see also Ashkun
- Nusu 968–969
see also Lolo-Burmese languages
- Nutka *see* Nuuchahnulth
- Nuuchahnulth (Nootka) 788–791, 1157
 augmentative 789
 classification 252
 compounding 1160
 coordination 790
 diminutives 789
 fixation 789
 head initial 789
 head marking 789
 morphemes 789
 morphology 789
 nominal phrase 1160
 noun incorporation 790
 phonology 788
 consonants 788, 789t
 delabialization 788–789
 glottalization 788–789
 labialization 788–789
 lenition 788–789
 vowel coalescence 788–789
 vowels 788
- predicates 789
 reduplication 789
 relative clauses 789–790
 resources 790
 suffixation 789
 syntax 789
 use of 788
 VSO 790
 word order 789
see also Wakashan languages
- Nyabwa 624
see also Guere languages
- Nyahkur 727
see also Monic languages
- Nyangumarda (Nyangumarta) 90
see also Australian languages
- Nyanja 791–797
 classification 253
 dialects 791
 dictionaries 794
 diminutives 796–796
 grammars (books) 794
 Henry, George 794
 Hetherwick, Alexander 794
 history/politics 791
 literary tradition 794
 name reversion 793
 noun classes 796–796
 nouns 794
 possessives 795–796
 regional politics 792
 Riddel, Alexander 794
 Scott, David 794
 as tonal language 794–795
 use of 791
 media 792
 verbs 794
see also Bantu languages; Kordofanian languages; Niger-Congo languages
- Nymylan *see* Koryak
- Nyo 771
see also Kwa languages
- Nyorsk, Norwegian 785
- Nzema 631
see also Tano languages

O

- Oasis dialects, North Arabian 932
- Oaxacan languages
 eccentric noun classes 823
 exocentric noun classes 823
 noun classes 823
- Obi-Urgic *see* Uralic languages
- Obligatory Contour Principle (OCP), Bantu languages 139–140
- Obo Manobo
 case marking 1005–1006, 1006t
 ergative 1005–1006
 ergative patterns 1005–1006, 1006t
 morphology 1003
 morphosyntax 1005
 pronouns 1005–1006, 1006t
 transitive sentences 1005–1006
see also South Philippine languages
- Ob-Urgic
 classification 1129–1130
 main verb phrases 1132
 negation 1131
 plural markers 1131
 word order 1132
- Occidental 77
- Occitan 799–800
 classification 251–252
 dialects 799
 diversity 799
 history 800
 texts 800
 official status 799
 perspectives 800
 revival/survival 800
 status 799
 structure 799

- Occitan (*continued*)
 Catalan *vs.* 799
 use of 799
see also Catalan; French; Romance languages
- Oceanic languages
 Austronesian languages 99
 classification 250–251, 685
see also Malayo-Polynesian languages
- Ofayé
 classification 665, 666, 666*t*, 667
 morphology 667
 speaker numbers 667
see also Macro-Jê languages
- Official Aramaic *see* Aramaic, Official
- Ofo 749
see also Siouan languages
- Ogham
 Gaelic script 200
 Goidelic languages 453
- Oghuz, South, related languages, Azerbaijani 110–111
- O'Grady, G N, Australian language
 classification 84
- Oirat 722
 use of 723
see also Mongol languages
- Ojibwa
 bilingualism 261
 classification 25
 dialects 755, 756*t*
 language shift 319
 origin/development 28
 phonology, stress 26
 speaker numbers 26
see also Algonquin languages
- Okere 631
see also Guang languages
- Okinawan *see* Ryukyuan
- Ok languages 1087
see also Trans New Guinea languages
- Oko 151
see also Benue-Congo languages
- Old Aramaic *see* Aramaic
- Old Church Slavonic 800–802
 classification 251–252
 Macedonian, influence on 663
see also Balkan linguistic area;
 Balto-Slavic languages; Bulgarian;
 Church Slavonic; Indo-European
 languages; Macedonian; Russian;
 Slavic languages; Sogdian
- Old Dutch (Old Low Franconian) 307
- Old English 356–359
 case marking 359
 dialects 356
 genetic relationships 356
 Old Frisian *vs.* 356–357
 Old Norse *vs.* 356–357
 Old Saxon *vs.* 356–357
 grammar 358
 inflectional morphology 358–359
 influence on other languages
 Old Icelandic 781
 Present Day English 329–330
 influences from other languages 358
 Celtic 358
 French 358
 Latin 358
 Scandinavian languages 358
 mutation 358–359
 origin/development 357
see also Germanic languages
 period used 356
 phonology 357
 fricatives 358
i-umlaut 357
 syllable structure 357
 verb positions 359
 written records 357
 Latin script 357
see also English, Early Modern; English, Later
 Modern; Germanic languages; Gothic;
 Middle English
- Old French Sign Language 956
- Old Georgian 291
- Old Icelandic *see* Icelandic, Old
- Old Irish *see* Irish, Old
- Old Javanese script, Balinese 117
- Old Kirghiz 1110
see also Turkic languages
- Old Norse *see* Norse, Old
- Old Persian *see* Persian, Old
- Old Saxon *see* Saxon, Old
- Old Southern Arabian 931
see also Semitic languages
- Old Uyghur 1110
see also Turkic languages
- Olmos, Andrés de, Nahuatl 745
- Olonetsian 1129–1130
see also Finnic languages
- Oluta, aspects 714
- Oluteco 713
see also Mixe-Zoquean languages
- Omaha-Ponca 802–805
 auxiliaries 804
 classification 252
 as active-stative language 803
 as endangered language 802
 as pronominal argument
 language 804
 dialects 804
 downstep 803
 gendered speech 804
 morphology 803
 orthography 802
 phonology 802, 802*t*
 nasality 803
 vowel length 803
 possessives 803
 relative clauses 804
 revival/survival 802
 SOV 804
 spelling 802
 subordinate clauses 804
 syntax 804
 use of 802
 USA 802
 verbs 803
 intransitive verbs 803
 transitive verbs 803–804
 word order 804
see also Crow; Siouan languages
- Oman, Arabic 42
- Ometo 805
see also Omotic languages
- Omotic languages 12, 273, 805–806
 classification 250
 definition 805
 as tonal language 806
 use of 805
see also Afroasiatic languages; Ari; Cushitic
 languages; Ethiopian linguistic area (ELA);
 Wolaitta
- Oneida 806–809
 affixation 806–807
 classification 252
 derivational process 806–807
 future 807
 genders 806–807
 history 806
 morphology 806
 nouns 807
 phonology 806
 vowels 806
 preservation/recovery 808
 scholarship 808
 syntax 807
 verbs 807
 word building 807
 word classes 806–807
 word order 807–808
 workers in, 808
see also Iroquoian languages
- Onomatopoeia, Korean 615–616
- Onondaga 543
see also Iroquoian languages
- Oosgrens Afrikaans 8, 9*f*
- Oowekyala 1157
 glottalized vowels 1158
see also Wakashan languages
- Opón-Carare *see* Cariban languages
- Oral tradition, Avestan 107
- Oranjeriver Afrikaans 8, 9*f*
- Oraon *see* Kurukh
- Ordinal derivation, European linguistic area 400, 401*f*
- Ordos 722
see also Mongol languages
- Orejón
 consonants 1095
 nasalization 1095
 speaker numbers 1092*t*
see also Tucanoan languages
- Oriya
 Assamese *vs.* 78
 classification 522
 compound verbs 996–997
 converbs 995
 number of speakers 523
 phonology 525–526
 vowels 526–527
see also Indo-Aryan languages
- Oromo 809–812
 consonant clusters 810
 dialects 809
 morphology 810
 as national language 809
 nominative case 810–811
 number of speakers 272–273
 person markers 811–812
 phonemes 809–810, 810*t*
 phonology 809
 political effects 809
 possessives 810–811
 SOV 812
 as tone-accent language 810
 use of 809
 verbs 811
 vowels 810
 word order 812
see also Afroasiatic languages; Cushitic
 languages; Ethiopian linguistic area (ELA)
- *Oryan Tepe 112–113
see also Azerbaijani
- OSP, Hiri Motu 501
- Ossetic 812–819
 adjectives 815
 agglutinative 814
 aktionsart 816, 817*t*
 aspect 816, 817*t*
 Caucasian languages, influence from 814
 classification 251–252
 copular sentences 818
 correlative 818
 declensions 540
 definite articles 540
 demonstratives 815
 dialects 813
 directional prefixes 542
 embedding 817*f*, 818, 818*f*
 enclitics 815, 816*t*
 ethnography 812
 future 817*t*
 genitives 815
 habitual 816
 habitual present 816, 817*t*
 history 812
 imperfective 816
 literature 812
 momentaneous present 816, 817*t*
 nouns 814
 noun phrases 816
 numerals 815, 816*t*
 objects 814–815
 phonology 540
 accent 814
 affricates 813
 consonants 540, 813, 814*t*
 loanwords 814
 postalveolar affricates 813
 vowels 814, 814*t*

- plurals 814–815, 815*t*
 postposition phrases 816
 pronouns 815, 815*t*
 simple sentences 817
 SOV 817
 tenses 815–816, 816*t*
 Tocharian, influence on 1070
 use of 812
 geographical distribution 813*f*
 verbs 815
 word order 817
see also Indo-European languages; Iranian languages
 Ostapirat, W, Austro-Tai hypothesis 105
 Ostayak-Samoyed *see* Selkup (Ostayak-Samoyed)
 Ostocharisch *see* Tocharian
 OSV, Thai 1059
 Oti 665–666, 666*t*
 see also Macro-Jê languages
 Oti-Volta languages 770
 see also Gur languages
 Oto-Mangean languages 748
 alignment 822
 locative adpositionals 822
 noun phrases 822
 verbs 822
 classification 819–821
 comparative phonology 819
 as endangered language 824
 Swadesh, Morris 819
 classifiers 823
 cliticization 822
 consonants 821
 constituent order 822
 documentation 823
 grammar 822
 historical aspects 751
 Hokan languages *vs.* 505
 noun classes 823
 endocentric noun classes 823
 origin/development 823
 phonology 821
 pronouns 822
 SOV 823–824
 syllabic nuclei 821
 syllable onsets 821
 morpheme patterns 821
 time depth 819
 tone 821
 use of 819
 speaker numbers 824
 viability 823
 vowels 821
 VSO 822
 workers in, Swadesh, Morris 819
 see also Amuzgo; Amuzgoan languages; Native American languages
 Otomian *see* Otomapean languages
 Otomí languages 819
 classification 819–821
 see also Oto-Mangean languages
 Oto-Palmean, time depth 819
 Otomapean languages 751
 see also Chichimeca Jonaz
 Oto-Panean-Chinanteko languages 819–821
 see also Oto-Mangean languages
 Oto-Panean languages
 classification 819–821
 see also Oto-Mangean languages
 Ottoman, contacts, Tatar 1053
 Otúke *see* Boróro (Otúke)
 Overlapping exponence, in isolating language 223
 OVS
 Arabic 49
 Cariban languages 186
 Luxembourgish 659–660
 Madang languages 671
 Spanish 884
 Trans New Guinea languages 1087
 Tucanoan languages 1096
Oxford History of the English Language 355
- P**
 Pacoh 726–727
 see also Katic languages
 Padang *see* Dinka
 Padari 526–527
 see also Indo-Aryan languages
 Padaung 581
 see also Karen languages
 Páez *see* Barbacoan languages
 Pahari
 classification 522
 phonology 526–527
 see also Indo-Aryan languages
 Pahlavi (Middle Persian) 827–828
 classification 251–252
 declensions 540
 eteograms 827
 heterograms 827
 influences from other languages 827
 literature 827
 past tenses 541
 pronouns 541
 script 827
 Zoroastrianism 538, 827
 see also Iranian languages; Persian, Modern; Persian, Old
 Paiwan
 classification 421
 dialects 421
 dictionaries 423
 grammars (books) 423
 research history 423
 see also Formosan languages
 Paiwanic languages 250–251
 see also Austronesian languages; Formosan languages
 Pajalat 506
 see also Hokan languages
 Pakatan 728–729
 see also Chut languages
 Pakistan
 Balochi 134, 538
 Brahui 162–163
 Burushaski 175
 Dardic languages 282, 283
 Hindi 495
 Indo-Aryan languages 522
 Iranian languages 537
 Kashmiri 582
 official language 885–886, 1133
 Pashto 538, 845
 Punjabi 885
 Shina languages 283
 Sindhi 960
 Tibetan 1060–1061
 Urdu 522–523, 885–886, 1133
 Pakistani Baluchistan 846
 Palaic
 features of 37
 historical aspects 36
 Palaihnihan languages 750–751
 see also Hokan languages
 Palare 548
 Palaung languages 727
 Shan, influence from 728
 use of 728
 varieties 728
 see also Palaung-Wa languages
 Palaung-Wa languages 724
 use of 727
 see also Angkuic languages; Mon-Khmer languages
 Palaychi 581
 see also Karen languages
 Palenquero 828–830
 classification 249*t*
 code switching 828–829
 development, isolation 828
 influence from other languages
 Kikongo 829
 Spanish 828
 lexicon 828
 resources 829–830
 use of 828
 Colombia 828
 geographical distribution 828*f*, 829*f*
 see also Creoles; Pidgins
 Pali 830–833
 alphabet 831
 canonical texts 830–831, 831–832
 classification 251–252
 commentaries 832
 correlative 831
 influence on other languages
 Khmer 248
 Lao 640
 Thai 248
 morphology 831
 non-canonical literature 832
 origin/development 830
 phonology 525, 831, 831*t*
 Sanskrit *vs.* 831
 script 831
 Thai, influence on 1059
 use of
 Burma/Myanmar 830
 Cambodia 830
 Sri Lanka 830
 Thailand 830
 word order 831
 see also Indo-Aryan languages; Indo-Iranian languages; Sanskrit; Tibetan
 Palikur
 classifiers 61
 genders 61
 predicate structure 60
 see also Arawak languages
 Palmella 185*f*
 see also Cariban languages
 Palula 282
 see also Shina languages
 Palyu 729
 see also Mon-Khmer languages
 Pama-Nyungan languages 84
 classification 250
 ergative 88
 morphology 87
 nouns 87
 pronouns 84–85, 87
 syntax 87
 see also Australian languages
 Pamean languages 751
 classification 819–821
 see also Oto-Mangean languages
 Pame languages 819–821
 see also Oto-Mangean languages
 ‘Pan-African properties,’ Africa, as linguistic area 4, 5*t*
 Panama
 Andean languages 40
 Chico languages 224
 Choco languages 40
 Emberá 224
 Waunmécú 224
 Panará
 ergative 668
 noun incorporation 667
 see also Jê languages
 Panare
 geographical distribution 185*f*
 morphemes 184–185
 phonology 183–184
 see also Cariban languages
 Panche *see* Cariban languages
 Panduro, Lorenzo Hervas, Austronesian languages 98
 Pangasinan 783
 see also North Philippine languages
 Panoan languages 750
 classification 252–253, 833
 de la Grasserie, Raoul 833
 ergative 834
 evidentiality 834
 history/culture 833
 lexicon/ethnolinguistics 834
 morphology 834
 phonology 834

- Panoan languages (*continued*)
 syntax 834
 use of 833
see also Native American languages
- Panzaleo 41
- Pa-O (Taungthu) 581
 writing systems 581
see also Karen languages
- Papantla Totonac 1080
 consonants 1082
 nouns 1083
 object agreement 1084
 speaker numbers 1081
see also Totonacan languages
- Papiamentu 835–836
 classification 249*t*
 habitual 835
 imperfective 835
 morphology 835
 origins 835
 orthography 835
 phonology 835
 progressive 835
 SVO 835
 syntax 835
 use of 835
see also Creoles; Dutch; Pidgins
- Papora-Hoanya 421
see also Formosan languages
- Papua New Guinea
 Cebuano 197
 English 836
 language density 319
 Manambu 693
 national languages 1076–1077
 Papuan languages 836
 Skou languages 973
 Tok Pisin 1076–1077
 Torricelli 1078
- Papuan languages 836–845
 classification 840
 Ross, Malcolm 837*f*
 dictionaries 841
 diversity 842
 geographic barriers 842–843
 isolation 843
 social/political organization 842–843
 existential verbs 842
 grammars (books) 841
 Island Melanesia 841
 morphology 841
 nominal inflexions 842
 North New Guinea languages 840
 noun classes 842
 noun roots 842
 phonetics 841
 pronominal systems 841
 research history 839
 Greenberg, Joseph H 839–840
 Ray, S H 839
 SOV 841
 SVO 841
 use of 836
 geographical distribution 837*f*
 New Guinea 836
 speaker numbers 836
 verbs 842
 verb roots 842
 word order 841
 workers in
 Greenberg, Joseph H 839–840
 Ray, S H 839
 Ross, Malcolm 837*f*
see also Austronesian languages; Central
 Solomons languages; Madang
 languages; Malayo-Polynesian
 languages; Malukan languages;
 Torricelli languages; Trans
 New Guinea languages; West Papuan
 languages
- Paraguay, Guaraní 467
- Paraujano *see* Arawak languages
- Parengi 996–997
- Parent languages, Nostratic theory 786–787
- Parsic *see* Judeo-Persian
- Parthian 538
 declensions 540
 Kurdish, influence on 625
 past tenses 541
 pronouns 541
see also Iranian languages
- Participial passive, Standard Average European
 (SAE) languages 393–394
- Particle comparatives, Standard Average
 European (SAE) languages 393–394
- Particles
 Anatolian languages 37
 Thai 1060
- Pashai languages
 agreement patterns 284
 case-marking 284
 classification 282
 sibilants 283
see also Dardic languages; Iranian
 languages
- Pashto 845–849
 ‘auxiliation,’ 847
 classification 251–252
 conjugation 847
 dialects 845, 845*t*
 Pakistani Baluchistan 846
 Waziri metaphony 845–846
 differential object marking 848
 dual 540
 ergativity 848
 imperfective 847
 landey 849, 849*t*
 morphology 847
 nouns 847
 order of terms 847, 848*t*
 origin/development 845
 orthography 846
 Arabic script 846
 Persian, influence from 846
 phonology 540, 846
 consonants 845, 845*t*, 846, 846*t*
 word-initial clusters 846, 847*t*
 present tenses 847
 pronouns 847, 847*t*
 SOV 848
 syntax 847
 use of 538, 845
 Afghanistan 538, 845
 Pakistan 538, 845
 verbs 847
 anti-impersonal verbs 848
see also Avestan; Balochi; Indo-Iranian
 languages; Iranian languages; Lahnda;
 Persian, Modern; Urdu
- Passive
 Arabic 51, 51*t*
 Bengali 150
 Indo-Iranian languages 534
 in introflexing languages 51, 51*t*
 Kurdish 626
 Norwegian 786
 Persian, Modern 851
- Patagonia
 Andean languages 41
 Mapuche 41
 Welsh 1169
- Patronymic names, Icelandic 782
- Pawnee 749
see also Caddoan languages
- Pazeh
 classification 421
 research history 423
see also Formosan languages
- Peano, Giuseppe, fields of work, artificial
 languages 77
- Pear 728
see also Pearic languages
- Pearic languages 724
 use of 728
see also Mon-Khmer languages
- Pederson, Holger, Nostratic theory 530
- Pehuenche 701
see also Mapudungan languages
- Pelliot, Paul
 Sogdian 986
 Tocharian 1068–1069
- Peloponnesian-Ionian Greek
 dialect 465
- Pemong
 gender 184–185
 geographical distribution 185*f*
 vowels 183–184
see also Cariban languages
- Pengo 300*t*
see also Dravidian languages
- Pennsylvania Dutch 444
- Penobscot 26
see also Algonquin languages
- Pentreath, Dolly 257–258
- Penutian languages 750
 historical aspects 747–748
see also Yokutsan
- Pequot-Mohegan-Montauk 24
see also Algonquin languages
- Perfect
 Arabic 51
 Azerbaijani 112
 Bactrian 115
 Balkan linguistic area 129, 130*t*
 Bengali 149
 Berber languages 157*t*
 Brahui 165
 Bulgarian 169
 Domari 296
 European linguistic area 393–394
 Finnish (Suomi) 414
 French 428
 Goidelic languages 453
 Greek, Ancient 463
 Greek, Modern 466
have 129
 Indo-Iranian languages 533–534
 Iranian languages 537
 Iroquoian languages 544
 Italian 546
 Kalkutungu 575
 Khotanese 604
 Krio 619
 Kru languages 624
 Latin 642
 Lithuanian 648
 Macedonian 664–665
 Persian, Modern 851
 Persian, Old 853
 Romani 900
 Romanian 902–903
 Slavic languages 977
 Slovak 979
 Sogdian 986
 Sorbian 994
 Southern Bantu languages 1017
 Spanish 1021
 Swedish 1030
 Syriac 1033
 Tajik 1042
 Xhosa 1192
 Yanito 1202
- Periphrasis *see* Clitic(s); Inflection
- Permian (Permian) languages
 classification 1129–1130
 phonology, consonantism 1130
see also Uralic languages
- Persian
 classification 251–252
 influence on other languages
 Azerbaijani 112
 Bengali 148
 Hindi 495–496
 Kashmiri 582–583
 Kurdish 625
 Malayalam 680–681
 Pashto 846
 Punjabi 889
 Urdu 1134
 Middle *see* Pahlavi (Middle Persian)
 Modern *see below*
 Old *see below*

- Urdu literature 1137
see also Iranian languages
- Persian, Modern 849–852
 classifiers 850
 definite markers 850
 dialects 850, 851
 future 851
 genders 540, 850
 grammars (books) 850
 indefinite markers 850
 indirect speech 851–852
 morphology 850
 origin/development 849
 passive constructions 851
 passives 851
 past continuous 851
 perfect 851
 phonology 850
 vowels 850
 pluperfect tense 851
 plurals 850
 possession 850
 preverbs 851
 relative clauses 851
 syntax 851
 tenses 541
 see also specific tenses
 use of 538
 Afghanistan 538, 850
 Iran 538, 850
 Tajikistan 850
 Tajikistan 538
 verbs 851
 workers in, Jones, William 850
 writing 538
 Arabic script 850
see also Arabic; Aramaic; Avestan; Bengali;
 Iranian Languages; Iranian languages;
 Kurdish; Pahlavi (Middle Persian); Pashto;
 Persian, Old; Punjabi; Tajik Persian;
 Türkmen
- Persian, Old 537, 852–854
 adjectives 853
 Avestan divergence 107
 correlative 853
 definition 852
 inflectional morphology 853
 inscriptions 852
 cuneiform script 852
 extent 852
 trilingual 852
 morphology 539
 nouns 853
 perfect 853
 periphrastic past tense 853
 phonology 852
 pronouns 853
 relative pronouns 853
 use of 852
 verbs 853
see also Akkadian; Avestan; Bengali;
 Elamite; Indo-European languages;
 Indo-Iranian languages; Iranian
 languages; Pahlavi (Middle Persian);
 Persian, Modern; Sanskrit; Sogdian;
 Tajik Persian
- Perso-Arabic
 Punjabi writing systems 886
 Telugu, influence on 1055, 1058
- Personal indicators, North
 American native language
 variation 756
- Person markers 811–812
- Peru
 Andean languages 41
 Arawak languages 59
 Aymara 108
 Aymaran 41
 Campa languages 59
 official languages 891
 Panoan languages 833
 Quechua 41, 891
 Tucanoan languages 1091
 Pharyngeal fricatives 380
- Philippines
 Austronesian languages 97, 99
 Cebuano 197
 English 1035
 Hiligaynon 492
 Ilocano 518
 official languages 1035
 Samar-Leyte 914
 Tagalog 783, 1035
 Pho 581
see also Karen languages
- Phoenician 854–855, 933
 alphabet 854
 Ammonite *vs.* 854
 classification 250
 Edomite *vs.* 854
 Hebrew *vs.* 854
 imperfective 855
 influence on other languages 854
 inscriptions 854
 Moabite *vs.* 854
 nouns 854
 use of 854
 verbs 855
 word order 855
see also Afroasiatic languages; Northwest
 Semitic languages; Semitic languages
- 'Phoenician shift,' 854–855
- Phonetics
 Assamese 78
 Nuristani languages 787
- Phonological form invariance, in isolating
 language 223
- Phonology, gender assignment 757
- Phunoi 968–969
see also Lolo-Burmese languages
- Pictish 855–857
 affiliation of 855–856
 classification 251–252
 as non Indo-European language 856
 origin/development 855
 Scotland 855
 workers in
 Camden, William 856
 Dunbavin, Paul 855–856
 Fraser, John 856
 Jackson, Kenneth H 856
 Johnston, J B 856
 Macalister, R A S 856
 Macbain, Alexander 856
 Nicolaisen, W F H 856
 Rhys, John 856
 Skene, W F 856
 Sverdrup, Harald V 856
see also Brythonic Celtic; Cornish; Finnish
 (Suomi); Indo-European languages; Scots
 Gaelic; Welsh
- Pictographs, Chinese development 217
- Picunche 701
see also Mapudungan languages
- Picuris 1049
see also Tanoan
- Pidgins 857–864
 Algonquin languages 28
 Assamese 78
 basilect 867
 classification 858
 lexical affiliation 858
 clicks 859
 Creoles *vs.* 862
 definition 859
 development 249
 education 867
 future research 863
 habitual 867–868
 imperfective 867–868
 lectal level 867
 Portuguese, influence from 883
 pro-drop 865
 progressive 867–868
 SVO 862
 tense-mood-aspect systems *vs.* 858
 types 249*t*
 variations in 865
- see also* African-American Vernacular
 English (AAVE); Austronesian languages;
 Bislama; Cape Verdean Creole; Creoles;
 English, nonnative; Fanagalo; Gullah;
 Hawaiian Creole English (HCE); Krio;
 Louisiana Creole; Mobilian Jargon
 (Mobilian); Morrobalama; Palenquero;
 Russenorsk; Sango; Tiwi; Tok Pisin;
 Yanito
- Pidgin Swahili 140
see also Bantu languages
- Pie 624
see also Grebo languages
- Piedmontese *see* Italian
- Pied-piping 1214
- Pijao *see* Cariban languages
- Pijin 836
- Pimenteria 185*f*
see also Cariban languages
- Pina, Francisco de 1149–1150
- Pinghua languages
 classification 969
 speaker numbers 214*t*
see also Chinese
- Pinto, Constancio 226
- Pipil, sound correspondences with
 Finnish 651
- Piratapuyo
 consonants 1094*t*
 speaker numbers 1092*t*
 syllable pattern 1095
see also Tucanoan languages
- Pisaflores Tepehua 1081
 consonants 1082
 inflectional affixes 1083
 reciprocal verbs 1083
 speaker numbers 1082
see also Totonacan languages
- Pisamira
 case markers 1097
 consonants 1094*t*
 morphemes 1096
 speaker numbers 1092*t*
see also Tucanoan languages
- Piscataway *see* Conoy (Piscataway)
- Pite Saami 911
see also Saami
- Pitjantjatjara 871–874
 ceremonial vocabularies 871
 complex sentences 873
 ergative 872
 example 873
 future 873*t*
 history 871
 imperfective 873
 morphology 90, 872
 nominal morphology 872
 ergative case allomorphy 872
 locative case morphology 872
 pronouns 872, 873*t*
 phonology 872
 consonants 872, 872*t*
 vowels 872
 sociolinguistics 871
 syntax 90, 872
 use of 871
 speaker numbers 871
 verbs 88, 872, 873*t*
 alternate forms 871
 serial verb constructions 871
 tense-aspect-mood
 categories 872–873
 word order 871
see also Australian languages; Pama-Nyungan
 languages; Warlpiri
- Pitta Pitta
 future 88
 morphology 87, 88
 syntax 87, 88
 verbs 88
see also Australian languages
- Pittman, Richard S, *Ethnologue* 385
- Platoid 151
see also Benue-Congo languages

- Plural(s), language diffusion 248
 Plural sweep, sign language 951, 952f
 Pnar 595
 Poguli 282
 see also Kashmiri languages
 Poland
 German 444
 Slovak 977
 Police Motu *see* Hiri Motu
 Polish 874–878
 classification 251–252, 974–975
 declensions 875–876
 dialects 877
 genders 875
 imperfective 876
 iterative 876
 lexicon 877
 morphology 875
 nouns 875
 origin/development 877
 orthography 874
 Latin alphabet 874–875
 phonology 875
 consonants 875
 diphthongs 975
 vowels 875
 word stress 875
 plurals 875
 pronouns 877
 syntax 876
 verbs 876
 word order 876
 written records 877
 Yiddish, influence on 1205
 see also Balto-Slavic languages; Belorussian;
 Slavic languages; Sorbian
 ‘Politeness shift,’ 684
 Politics, European linguistic area 390
 Polivanov, E. D. 31–32
 Polyfunctional suffixes 419
 Polyglotta Africana 472
 Polymorphemic signs 952, 952f
 Polysynthesis, Native American languages 746
 Polysynthetic languages 201–204
 productive noninflectional concatenation *vs.*
 nonproductive morphology 204
 see also Algonquin languages; Arabic, as
 introflecting language; Caddoan
 languages; Central Siberian Yupik; Crow;
 Eskimo-Aleut languages; Lakota;
 Morphological Types; Nahuatl; Ngan’gi;
 Ritwan languages; Tiwi
 Pomoan languages 878–883
 as agglutinative languages 881
 assertion evidence 881
 classification 880f
 Barrett, Samuel A. 878
 clauses 881
 grammars (books) 878–879
 historical relationships 881
 intermarriage 878–879
 instrumental prefixes 881
 interrelationships 880f
 kinship terms 881
 morphology 881
 phonology 880
 consonants 880
 voiced stops 881
 voiceless aspirated stops 880–881
 pronouns 881
 as stative-active languages 881
 use of 878, 879f
 see also Hokan languages
 Pomo languages 750–751
 classification 505
 see also Hokan languages
 Pong 728–729
 see also Cuoi languages; Muong languages
 Pontic dialect 465
 Popoloca 751, 819
 see also Oto-Mangean languages; Popolocan
 languages
 Popolocan languages 751
 see also Chocho
 Popti’ 705–706
 noun classifiers 708
 numeral classifiers 709t
 speaker numbers 706t
 see also Mayan languages
 Poqom 705–706
 see also Mayan languages
 Poqomam 705–706
 speaker numbers 706t
 transitive verbs 708t
 see also Mayan languages
 Poqomchi’ 705–706
 speaker numbers 706t
 see also Mayan languages
 Portugal
 official languages 883
 Portuguese 883
 Portuguese 883–885
 adjectives 884
 characteristics 883
 classification 251–252
 Galician *vs.* 435
 history 883
 earliest texts 883
 literary texts 883
 influence on other languages
 Cape Verdean Creole 182
 Hindi 495–496
 Korean 615–616
 Krio 620
 Malayalam 680–681
 Saramaccan 858
 Sinhala 964
 Sranan 858
 Telugu 1058
 influences from other languages
 Galician 883
 Latin 883
 morphology 884
 phonology 884
 consonants 884
 diphthongs 884
 monophthongs 884
 SVO 884
 syntax 884
 use of 883
 Brazil 883, 884
 varieties 884
 Creoles 883
 pidgins 883
 verbs 884
 see also Catalan; Galician; Romance languages;
 Spanish
 Possessive doubling, Balkan linguistic area 125
 Possessives
 African-American Vernacular
 English (AAVE) 336
 Akan 19
 Austronesian languages 101
 Balkan linguistic area 124
 Balto-Slavic languages 136
 Cariban languages 186
 Chorasmian 238
 Chuvash 244–245
 Creek 264
 Danish 280
 Domari 296
 Egyptian 39
 English 340
 Ethiopian linguistic area (ELA) 380
 Ethiopian Semitic languages 383
 Evenki 406
 Finnish (Suomi) 416t
 Hausa 479
 Hokan languages 509
 Kayardild 585–586
 Kazakh 590
 Kinyarwanda 608
 Kirghiz 612
 Krio 617
 Munda languages 737
 Nenets (Yurak) 762
 Nyanja 795–796
 Omaha-Ponca 803
 Oromo 810–811
 Santali 922–923
 Sumerian 1024
 Swahili 1027t
 Swedish 1031
 Tatar 1054
 Tohono O’odham 1075
 Turkic languages 1112
 Turkish 1116
 Uyghur 1144
 Wambaya 1163
 Wolaitta 1181–1182
 Xhosa 1188
 Yakut 1200
 Postpositions
 Ainu 17
 Akan 19
 Balkan linguistic area 122
 Bengali 149
 Burmese 173
 Cariban languages 186
 Caucasian languages 195
 Creek 741
 Crow 269
 Cushitic languages 274–275
 Dravidian languages 301
 Ethiopian linguistic area (ELA) 380
 Gondi 456
 Hindi 496–497
 Hiri Motu 501
 Hungarian 290
 Kashmiri 583–584
 Korean 616
 Kurukh 628
 Lakota 638
 morphological types 733–734
 Muskogean languages 741
 Navajo 761
 Nenets (Yurak) 762
 Ossetic 816
 Punjabi 888
 Sindhi 962
 Sinhala 965–966
 Siouan languages 972
 Tajik Persian 1042
 Tamil 1048–1049
 Telugu 1056
 Tupian languages 1106
 Turkish 1113–1114
 Uralic languages 1131–1132
 Votic 290
 Potawatomi
 classification 25
 speaker numbers 26
 see also Algonquin languages
 Poutsma, Hendrik, Later Modern English
 definition 343
 Powadi Punjabi 886
 see also Punjabi
 Powell, John Wesley, Uto-Aztecan
 languages 1140
 Powhatan
 classification 24
 origin/development 28
 see also Algonquin languages
 Pragmatics, Southeast Asian languages 1011
 Prakrit, influence on other languages,
 Telugu 1058
 Prasun 787
 see also Nuristani languages
 Prefixes
 diachronic origins 287
 suffixation *vs.* 288
 Pre-Proto-Indo-European (PPIE) 530
 Prešeren, France 981
 Principe Islands 883
 Pro-drop
 Czech 277–277
 Hausa 479
 Hungarian 516
 Italian 547
 Norse 781
 Pidgins 865
 Punjabi 889

- Telugu 1058
 Turkish 1116
 Productive noninflectional concatenation (PNC) 203
 Productivity, postbases 202–203
 Progressive
 Berber 156*t*
 Bikol 159*t*
 Breton 167
 Chinantecan languages 212*t*
 English 345–346
 Guugu Yimithirr 474–475
 Highland East Cushitic (HEC) languages 490
 Hiligaynon 493*t*
 Krio 619
 Kwa languages 632
 Papiamentu 835
 Pidgins 867–868
 Sogdian 537–538
 Somali 987
 Tajik Persian 1042
 Tucanoan languages 1100
 Turkish 1114–1115
 Waray-Waray 915*t*
 Xhosa 1191
 Pronominal(s)
 clitics *see* Clitic(s)
 object doubling 124
 subject markers, in introflecting language 52, 52*t*
 Pronouncing dictionaries, Later Modern English 344, 347
 Proom (Souei) 726–727
 see also Katuic languages
 Proper names, Hurrian 516
 Prosodic nasality (nasal harmony), Guaraní 468
 Proto-Algonquian
 classification 24–25
 consonants 27*t*
 reconstruction 26
 stress 26
 see also Algonquin languages
 Proto-Austronesian 684
 phonology 684
 see also Malayo-Polynesian languages
 Proto-Bantu *see* Bantu
 Proto-Central-Algonquian 651
 Proto Central/Eastern Malayo-Polynesian (PCEMP)
 classification 685
 evidence for 685–687
 see also Malayo-Polynesian languages
 Proto-Dravidian
 consonants 298, 298*t*
 vowels 297–298, 298*t*
 Proto-Germanic 447–448
 First Sound Shift (Grimm's Law) 448*t*
 Verner's Law 449*t*
 see also Germanic languages
 Proto-Indo European (PIE)
 Brugmann, K 529
 definition 530
 phonology, Tocharian *vs.* 1069
 see also Indo-European languages
 Proto-Je 651
 Proto Oceanic
 evidence for 687
 use of 687
 see also Malayo-Polynesian languages
 Proto-Semitic 929
 see also Semitic languages
 Proto-Sino-Tibetan
 morphology 970
 prefixes 970
 suffixes 970
 syllable structure 969–970
 use of 968
 see also Sino-Tibetan languages
 Proto Trans New Guinea language
 phonemes 1086*t*
 pronouns 1086*t*
 use of 1087
 Proto-Uralic *see* Altaic languages; Areal linguistics; Uralic languages
 Proto-World, long-range comparisons 649
 Provençal *see* Occitan
 Puare (Puari)
 subject marking 974
 vowels 973
 see also Skou languages
 Pukapuka, Niuean, influence on 776
 Pulaar *see* Fula
 Pumi
 classification 968–969
 see also Qiangic languages
 Punjabi 885–889
 adjectives 887–888
 cases 887
 classification 251–252, 522
 dialects 886
 future 888
 genders 887
 history 886
 influence from other languages 889
 Kashmiri, influence on 582–583
 literature 886
 morphology 887
 negatives 889
 nouns 887, 888*t*
 numbers 887
 phonology 525–526, 886
 aspirated *vs.* unaspirated consonants 886–887
 consonants 887, 887*t*
 germinates 887
 retroflexion 887
 tonal contrasts 886
 tonal system 525–526, 526–527
 vowels 526–527, 887, 887*t*
 postpositions 888
 prefixes 887
 pro-drop 889
 pronoun case system 888, 888*t*
 SOV 889
 stress 887
 suffixes 887
 syntax 889
 use of 885
 speaker numbers 523, 885–886
 varieties 886
 religions 886
 verbs 888
 causative verbs 888
 complex verbs 888
 compounding 888
 conjunct verbs 888
 prefixes 888
 stative/active distinctions 888
 tenses 888
 volitional/nonvolitional distinctions 888
 word order 889
 writing systems 524, 886
 Devanagari 886
 Gurmukhi 886
 Perso-Arabic 886
 see also Arabic; Dardic languages; Hindi; Indo-Aryan languages; Kashmiri; Lahnda; Persian, Modern; Sanskrit
 Puquina 752
 see also Native American languages
 Purí (Coroado)
 classification 665, 666, 666*t*
 geographical distribution 666–667
 records of 667
 see also Macro-Jê languages
 Puruborá 1106*t*
 see also Tupian languages
 Puruhá-Cañar 41
 Putonghua 215
 development 215
 Fuzhou *vs.* 219
 phonology 220*t*
 finals 215
 initials 215
 tones 215
 Puyuma
 classification 421
 dialects 421
 research history 423
 see also Formosan languages
 Pwo 968–969
 see also Karenic languages

Q

- Qafar *see* Afar (Qafar)
 Q'anjob'al 705–706
 noun classifiers 708
 speaker numbers 706*t*
 see also Mayan languages
 Qashqay 1112
 Qatabanian 931
 see also Semitic languages
 Qawaskar *see* Kawésqar (Qawaskar)
 Q'eqchi' 705–706
 speaker numbers 706*t*
 see also Mayan languages
 Qiangic languages 968–969
 see also Sino-Tibetan languages
 Qualifiers, Ijò 518
 Quapaw 749
 see also Siouan languages
 Quapaw, and Amerind 653
 Quechua languages 752
 agglutinating structure 892
 Aymara *vs.* 891
 classification 252–253
 dialects 891–892
 evidentiality 892–892
 grammar 891
 Mapudungan languages, influences on 702
 morphology 892
 nominalization 892
 nouns 892
 as official language 891
 phonology 892
 reconstruction 891
 stops 892
 subdivisions 891–892
 use of 891
 Bolivia 41
 Ecuador 41
 Peru 41, 891
 verbs 892
 vowels 892
 see also Andean languages; Aymara; Native American languages
 Question words, sign languages 958
 Quetzaltepec *see* Mixe-Zoquean languages
 Quileute 210
 morphology 210
 phonology 210
 typology 210
 voiced stops 211
 Quinault 749
 see also Salishan languages
 Quingnam 41
 Quirupí-Unquachog 24
 see also Algonquin languages
 Quotatives, South Asian languages 997
 Qur'an
 Sindhi translations 961
 see also Arabic, Classical; Islam

R

- Rabaul 444
 Rabbinic Hebrew 483
 Rabha 968–969
 see also Bodo-Koch languages
 Racism, Afroasiatic language investigations 12
 Rajasthani 525–526
 see also Indo-Aryan languages
 Ramarâma
 adjectives 1106
 ideophones 1106–1107
 phonology, tone system 1106
 Ramo 974
 see also Skou languages

- Ramstedt, Gustaf John, Altaic 31
 Ranquel 701
see also Mapudungan languages
 Rapoisi *see* Kunua (Rapoisi)
 Rashad 770
see also Kordofanian
 Rask, Rasmus Kristian
 Eskimo-Aleut 371
 Rathī Punjabi 886
see also Punjabi
 Ray, S H, Papuan languages 839
 Reading *see* Writing/written language
 Real character 76
 Recursion/iteration
 postbases 203
 sign language, syntax 944
 Red Tai (Tai Daeng) 1039
see also Tai languages
 Reduplication
 Apalai 184
 Balkan linguistic area 124
 Bislama 162
 Cape Verdean Creole 182
 Cariban languages 184
 Central Solomons languages 205
 Chadic languages 207
 European linguistic area 401, 402*f*
 Gikuyu (Kikuyu) 451–452
 Hausa 478
 Hindi 497
 Kinyarwanda 607
 Krio 621
 Madurese 673
 Nuuchahnulth (Nootka) 789
 Salishan languages 913
 South Asian languages 998
 Tamambo 1046
 Tiriyō 184
 Wayana 184
 Reference, sign language
 morphology 942, 943*f*
 Referential opacity, word *see* Word
 Referent tracking, converbs *see* Converbs
 Regional accents/dialects
 English 333
 North American native language
 variation 754
 Thai 1058
 Register tone, Kru languages 624
 Rek *see* Dinka
 Reland, Hadrian 98
 Relational concepts, language
 classification 731
 Relexification, Creole
 origins/development 860
 Religion, Austronesian languages 97
 Rembarnga (Rembarunga) 90
see also Australian languages
 Remo 736
see also Munda languages
 Replication, Balkan linguistic area 124
 Rere 613
see also Kordofanian languages
 Resígaro
 pronominal suffix loss 60
 tones 60
see also Arawak languages
 Resultative
 African-American Vernacular English (AAVE)
 336
 Balkan linguistic area 126
 Berber languages 156
 Bulgarian 169
 Creek 264–265
 Evenki 406
 Greek, Ancient 463
 Macedonian 664–665
 Mapudungan languages 702
 Slovak 975
 Resumptive clitic compounds, Balkan
 linguistic area 124
 Retroflex
 Burushaski 176
 Dardic languages 283–284
 Khotanese 603
 Punjabi 887
 South Asian languages 995
 Retuara/Tanimuca
 accent/tone 1096
 adjectives 1099
 case markers 1096
 classification 1091
 consonants 1094
 evidentiality 1100
 morphemes 1096
 nasalization 1095–1096
 noun classifiers 1098
 speaker numbers 1092*t*
 syllable pattern 1095
see also Tucanoan languages
 Rhaeto-Romance languages 893–895
 classification 251–252, 893
 phonology 894–895
 use of 893
 geographical distribution 893
see also Indo-European languages; Romance
 languages
 Rhodes, Alexander de, Vietnamese 1149–1150
 Rhys, John 856
 Riāng 727
see also Palaung-Wa languages
 Riau Indonesian 895–896
 Bazaar Malay *vs.* 895–896
 Riau Malay *vs.* 895–896
 Standard Indonesian *vs.* 895–896
 use of 895
see also Austronesian languages;
 Malay
 Riau Malay, Riau Indonesian *vs.* 895–896
 Riddel, Alexander 794
 Ridley, William 438
 Riis, H R 17–18
 Rikbaktsá
 classification 665, 666, 666*t*, 667
 use of 666–667
see also Macro-Jê languages
 Ringe, D A Jr. 655
 Ritwan languages
 classification 25, 252
 classifiers 28
see also Algonquin languages; Central
 Siberian Yupik; Cree; Michif;
 Mobilian Jargon (Mobilier); Polysynthetic
 languages
 Ro 75
 Roanoke-Pamlico 24
see also Algonquin languages
 Rokytno *Bund* 392–393
 Roman alphabet *see* Latin (Roman)
 alphabet
 Romance languages 896–898
 classification 251–252
 definition 896
 diminutives 896–897
 divergence 897
 influence on other languages
 Esperanto 376
 Spanish 1020
 Medieval speakers 897
 reconstruction 896
 texts 897
see also Catalan; French; Galician; Italian; Italic
 languages; Latin; Occitan; Portuguese;
 Rhaeto-Romance languages; Romanian;
 Spanish
 Romani 120, 898–900
 classification 251–252
 clauses 900
 definition 898
 diversity 900
 Domari *vs.* 295
 history 898
 influence from other languages 898–899
 morphology 899
 nouns 899
 adjectives 899
 genders 899
 ikeoclitc declension classes 899, 899*t*
 numbers 899
 phonology 899
 consonants 899
 dental clusters 898
 stops 899
 vowels 899
 syntax 900
 use of 898
 verbs 898, 899
 default stems 899–900
 perfective tense 900
 personal conjugations 900
 present tense 900
 valency 899
 word order 900
see also Armenian languages; Balkan
 linguistic area; Dardic languages;
 Domari; Dravidian languages; Hindi;
 Indo-Aryan languages; Iranian languages;
 Kashmiri
 Romania
 Gagauz 1112
 German 444
 Hungarian 514
 Romanian 901
 Slovak 977
 Romanian 900–904
 alphabet 901
 Balkan Sprachbund 902
 cases 902, 903*t*
 concord 900
 definite articles 902*t*
 dialects/dialectology 901
 earliest texts 901
 influences from other languages
 Latin 900–901
 Slavic languages 901
 morphology 902, 902*t*
 origin/development 900
 perfect 902–903
 phonetics 901
 consonants 901, 901*t*
 diphthongs 901
 palatal velars 901
 vowels 901*t*
 word stress 901–902
 Romance languages *vs.* 902
 tenses 902–903, 903*t*
 use of 901
 verbs 902–903, 902*t*
see also Balkan linguistic area; Romance
 languages
 Roman script *see* Latin (Roman) alphabet
 Romansh
 classification 893
 as national language 893
 use of 893
 geographical distribution 893
 speaker numbers 893
 Switzerland 893
see also Rhaeto-Romance languages
 Romany *see* Romani
 Ronga 1018
see also Tshwa
 Roots, affixes *vs.*, agglutinating *vs.* fusional
 languages 554
 Ross, Malcolm, Papuan language classification
 837*f*
 Roth, W E 473–474
 Rotokas 841
see also North Bougainville
 languages
 Rounded vowels, European linguistic area 395,
 396*f*
 Rounding harmony, vowels *see* Vowel
 harmony
 Ruc 728–729
see also Chut languages
 Rudhari 526–527
 Ruhlen, M 649
 Rukai
 classification 421
 dialects 421
see also Formosan languages

- Russell, Bertrand *see* Reference
- Russenorsk 904–905
classification 249*t*
lexicon 858
Norwegian 904
SVO 904
written material 904
see also Creoles; Pidgins
- Russia
Armenian 68
Caucasian languages 192
Evenki 405
Iranian languages 537
Kazakh 588
Ket 593
Kirghiz 610
Lak 636
Ossetic 812
Saami 911
Uralic languages 1129–1130
- Russian 905–908
Belorussian *vs.* 147
Church Slavonic features 905
classification 251–252, 974–975
declensions 976–977
future 907
grammar 906
grammars (books) 905
imperfective 907
influence on other languages 908
Abkhaz 2
Aleut 373
Azerbaijani 110–111, 112
Inuit 373
Inupiaq 536
Russenorsk 904
Tajik Persian 1043
lexicon 907
word borrowing 907
location, Uzbek bilingualism 1145
Nenets (Yurak) bilingualism 761–762
nouns 906
articles 907
feminine declension 906
masculine declension 906
predicative instrumental
standard 906
phonetics 905
accent 906
allophones 906
consonants 905
nonpalatalized consonants 906
palatalized consonants 906
voiced consonants 906
vowels 905
sound correspondences and comparisons
650–651
Ukranian *vs.* 1122–1123
use of
Khazakhstan 588
as official language 588
Uzbekistan 1145
verbs 907
Aktionsart 907
participles 907
tenses 907
'verbs of motion,' 907
written language 905
18th Century 905
diglossia 905
see also Balto-Slavic languages;
Belorussian; Old Church Slavonic;
Slavic languages; Tajik Persian;
Turkmen; Ukrainian
- Russian Federation, Buryat 723
- Rwanda
Kinyarwanda 604
national language 604
Swahili 1026
- Ryukyuan 908–909
Chamberlain, Alexander 908
classification 249
dialects 908
diminutives 907
grammar 908
origin/development 908
related languages, Japanese 557
speaker numbers 908
see also Altaic languages; Japanese
- S**
- Saami 911–912
books 911
classification 911, 1129–1130
as nominative-accusative
language 911
dialects 911
finite verbs 911
history 911
negation 911
nouns 912
phonology 912
consonant gradation 1130
consonantism 1130
consonants 912
diphthongs 1130
vowels 912
word stress 912
plural markers 1131
SOV 911
use of 911
word order 1132
see also Akkala Saami; Finno-Ugric languages;
Uralic languages
- Saami, South 911
see also Saami
- Saaroa
classification 421
research history 423
see also Formosan languages
- Sabaeen 931
see also Semitic languages
- Sabdamanidarpa 577
- Sach 728–729
see also Chut languages
- Sadani 526–527
see also Indo-Aryan languages
- Safaitic 932
see also Semitic languages
- Safe languages
definition 319–320
see also Language endangerment
- Sagart, L. 106
- Sahaptin-Nez Perce 750
see also Penutian languages
- Saharan languages
classification 773
geographical distribution 775*f*
see also Nilo-Saharan languages
- Sairaki 526–527
see also Indo-Aryan languages
- Saisiyat 421
see also Formosan languages
- Sakapulteko 705–706
speaker numbers 706*t*
see also Mayan languages
- Salina languages 504
classification 506
verbs 508
see also Hokan languages
- Salinan 750–751
see also Hokan languages
- Salishan languages 749
consonants 913, 913*t*
dictionaries 913–914
diminutives 913
as endangered language 913–914
genetic links 913
grammars (books) 913–914
morphology 913
phonology 913
reduplication 913
syntax 913
types 912–913
word classes 913
see also Bella Coola; Tillamook
- Sama, Southern
morphosyntax 1007
nouns 1007–1008
personal names 1007–1008
pronouns 1007–1008, 1008*t*
- Sama languages
classification 1002*t*
phonology 1002
see also Malayo-Polynesian languages
- Samar-Leyte 914–917
case markers 916, 916*t*
demonstratives 916, 916*t*
dialects 914
dictionaries 915
distinguishing features 915
lexicon 916
phonology 915
pronouns 915–916, 916*t*
related languages 914
Tagalog *vs.* 916
use of 914
Philippines 914
verbs 915–916, 915*t*
VSO 913
written works 915
see also Austronesian languages; Bikol;
Cebuano; Hiligaynon; Malayo-Polynesian
languages; Tagalog
- Samberigi (Sau) 1086–1087
see also Engan languages
- Samoa, Austronesian languages 102
- Samoan
classification 250–251
Niuean, influence on 776
possessive forms 101
see also Austronesian languages; Oceanic
languages
- Samoyed languages
classification 1129–1130
plural markers 1131
word order 1132
see also Uralic languages
- Samre 728
see also Pearic languages
- Sanchez, Mateo 915
- Sandawe 252
see also Khoesaa languages
- Sangali 613
see also Kadugli languages
- Sango 917–918
classification 249*t*
French, influence from 917
as official language, Central African Republic
917
origin/development 917
speaker numbers 917
see also Creoles; Pidgins
- San Marino, Italian 545
- Sanskrit 918–921
Avestan *vs.* 918–919
characteristics 919
classification 251–252
Indian cultural effects 920
influence on other languages
Bengali 148
Dravidian languages 306, 920–921
Hindi 495–496
Kashmiri 582–583
Khmer (Cambodian) 600
Lao 640
Malayalam 680–681
Punjabi 889
Telugu 1055, 1058
Thai 1059
morphology 919–920
origin/development 918
earliest records 919
standardization 919
Pani *vs.* 831
phonology 525, 920
present tense 533
religious influences 920–921
sample sentence 920
scripts, Devanagari 524

- Sanskrit (*continued*)
workers in 921
see also Avestan; Balkan linguistic area; Bengali;
Indo-Aryan languages; Indo-Iranian
languages; Pāli; Persian, Old; Punjabi;
Tocharian
- Santa 723
see also Mongol languages
- Santali 736, 921–924
classification 250
consonants 921
converbs 995
demonstrative system 922
dialects 921
morphology 737
possessives 922–923
use of 921
India 921
verbs 922
compound verbs 996–997
vowels 921
writing system 922
see also Austroasiatic languages; Munda
languages
- Sa'o'ch 728
see also Pearic languages
- São Tomé and Príncipe, Portuguese 883
- Sapir, Edward
morphological types 730
criticism of 731
Na–Dene languages 743
Uto-Aztec languages 1140
Wolof 1184
- Sapir-Whorf Hypothesis, artificial languages 76
- Sapoteko 819
classification 819–821
time depth 819
see also Oto-Mangean languages
- Sarab 112–113
see also Azerbaijani
- Saramaccan
influence from other languages 858
lexicon 858
- Sardinian 251–252
see also Romance languages
- Sassetti, Filippo 921
- Sau (Samberigi) 1086–1087
see also Engan languages
- Savara (Sora) 736
see also Munda languages
- Savosavo 841
classification 204
gender 205
use of 204
see also Central Solomons languages
- Saxon, Old
Old English *vs.* 356–357
Old Icelandic, influence on 781
- Sayuleño
auxiliary constructions 715
cliticization 714
inversive person marking 714–715
nouns 714
phonology 713
positional references 715
see also Mixe-Zoquean languages
- 'Scaldic' poetry, Old Icelandic 779–780
- Scandinavian languages
classification 251–252
influence on other languages
Danish 279
Old English 358
see also Germanic languages
- Schlegel, August Wilhelm von, morphological
types 730
- Schlegel, Friedrich von, morphological
types 730
- Schleicher, August
Indo-European languages 528–529
Lithuanian grammar 648
- Schleyer, Johan Martin 76
- Schmidt, Isaac-Jacob 723
- Schmidt, Johannes, Indo-European
languages 529
- Schmidt, P Wilhelm
Austic hypothesis 92
Malayo-Polynesian languages 684
- Schulze, Leonhardt, Khoesaa language
classification 601
- Schwa, stressed 122
- Scientific classification, Later Modern English
development 349–350
- Scotland, Pictish 855
- Scots 924–926
classification 251–252
concord 925
definition 924
English *vs.* 925
grammar 925
'Great Vowel Shift,' 924–925
origin/development 924
orthography 925
phonology 925
vowels 926
as recognized language 924
revival 925
Scottish Vowel-Length Rule 925–926
spelling 924
vocabulary 926
vowel length 925
written records 925
see also Dutch; Germanic languages; Scots
Gaelic
- Scots Gaelic 926–929
alphabet 927
characteristics 927
classification 200, 251–252
decline of 928
development 454
dialects 927
English loanwords 927
origins/development 926
pre-aspiration 927
revival 928
teaching in schools 928
sample 454
script, Ogham 200
speaker numbers 928
spelling 927
verbs 927
vowels 927
VSO 927
word order 927
see also Celtic; Goidelic Celtic; Goidelic
languages; Pictish; Scots; Welsh
- Scott, David, Nyanja 794
- Scottish Vowel-Length Rule 925–926
- Sebuano 99
see also Austronesian languages
- Second Sound Shift, German 445, 445*t*
- Secoya
accent/tone 1096
adjectives 1099
consonants 1095
evidentiality 1100
noun classifiers 1098
speaker numbers 1092*t*
syllable pattern 1095
verbs, evidentiality 1100
see also Tucanoan languages
- Secret languages, Australian
languages 91
- Sedang 726
see also Bahnaric languages
- Seediq
classification 421
dialects 421
research history 422–423
see also Atayalic languages
- Selkup (Ostyak-Samoyed)
case suffixes 1131
classification 1129–1130
negation 1131
verbs 1131
see also Samoyed languages
- Semantography 76
- Sembla 697–698
see also Mande languages
- Seminole/Creek 738–739
definition 263
see also Creek
- Semitic languages 12, 929–935
Central 931
classification 250
Nostratic theory 249
earliest record 12
East 930
influences on other languages
Pahlavi (Middle Persian) 827
Yiddish 1205
Northwest *see* Northwest Semitic languages
sentence structure 930
use of 929
West 930
alphabet 1121
see also Afroasiatic languages; Akkadian;
Amharic; Arabic; Arabic, as introflecting
language; Arabic, Classical; Arabic,
Middle; Arabic, North; Arabic, Southern;
Aramaic; Eblaite; Ethiopian Semitic
languages; Ge'ez; Hebrew, Israeli; Hebrew,
Pre-Modern, Biblical; Phoenician;
Sumerian; Syriac; Tigrinya; Yiddish
- Seneca 543
see also Iroquoian languages
- Senegal
Fulfulde 430
Mande 769–770
Mande languages 694
Wolof 1184
- Sentence(s)
Hiri Motu 501
Native American languages 746
structure 988, 989
- Senufo languages 770
see also Gur languages
- Sepedi 1187
- Sepik-Ramu languages
classification 253
nominal inflexions 842
noun classes 842
use of 840
see also Papuan languages
- Sera 770
see also Atlantic Congo languages
- Serbia and Montenegro
Albanian 22
Hungarian 514
- Serbian 936
- Serbian-Croatian-Bosnian Linguistic Complex
935–938
classification 251–252, 974–975
cultural divisions 936
dialects 936
morphology 936
orthography 937
use of, Italy 545
word order 936
see also Balkan linguistic area; Slavic languages;
Slovene
- Serbo-Croat *see* Serbian-Croatian-Bosnian
Linguistic Complex
- Sere languages 3
see also Adamawa-Ubangi languages
- Serial verbs
Central Solomons languages 205
Khmer (Cambodian) 725
Kinyarwanda 609
Krio 619
Mon-Khmer languages 725
Pitjantjatjara 871
Ravüa 725
Tamambo 1046
Tariana 1051
Torricelli languages 1079
- Seri languages 504
classification 506
person markers 507
see also Hokan languages
- Sesotho 1017
South Africa 1187
see also Sotho-Tswana languages

- Setswana *see* Tswana (Setswana)
 Seward Peninsula Inupiaq (SPI) 535
see also Inupiaq
 Sgaw 581
 classification 968–969
 writing systems 581
see also Karenic languages
 Shan
 classification 1039
 influence on other languages
 Palaung languages 728
 Wa 1156
see also Tai languages
 Shanghai Chinese 219
 development 219
 phonology 219
 use of 219
see also Chinese
 Shanxi 969
see also Jin languages
 Sharada, Kashmiri 583
 Sharchopkha *see* Tshangla
 (Sharchopkha)
 Shastan languages 750–751
see also Hokan languages
 Shawnee
 classification 25
 speaker numbers 26
see also Algonquin languages
 Sherbro 620
 Shevoroshkin, V 649
 Shina languages
 Burushaski, influence on 179
 case-marking 284
 classification 282
 phonology
 sibilants 283
 tonal system 526–527
 use of 283
 writing systems 524
see also Astor languages; Dardic languages;
 Indo-Aryan languages
 Shirumba 613
see also Kordofanian languages
 Shixing 968–969
see also Qiangic languages
 Shompeng 94
see also Austroasiatic languages
 Shona languages 1017
 classification 1017
 as tone language 938
 consonants 938
 dialects 938
 dictionary 938
 diminutives 938
 ideophones 938
 morphology 938
 nouns 938
 noun classes 938
 syntax 939
 use of 938
 Botswana 1017
 Mozambique 1017
 as official language 1017
 Zimbabwe 938, 1017
 verbs 938–939
 vowels 938
see also Bantu languages; Bantu languages,
 Southern
 Shoshonean 1139
see also Uto-Aztecan languages
 Shumashti
 agreement patterns 284
 classification 282
 phonology, sibilants 283
see also Kunar languages
 Shusha 112
see also Azerbaijani
 Shuswap 749
see also Salishan languages
 Siam *see* Thailand
 Siberia
 Nivkh 777–778
 Tungusic languages 1103
 Sichuan Yi 969
see also Hakka languages
 Sicily 464
 Sidaama
 noun morphology 491, 491*t*
 phonology 490–491, 490*t*
 use of 488–489, 488*t*
 verb morphology 490
see also Highland East Cushitic (HEC)
 languages
 Sidamo
 Ethiopia 272–273
 number of speakers 272–273
see also Cushitic languages
 Sieg, Emil 1068–1069
 Siegling, Wilhelm 1068–1069
 Sierra Leone
 Krio 617
 Mande languages 769–770
 Sierra Popoluca
 as agglutinative languages 714
 glottal stop metathesis 713
 persons 714
 transitive verbs 714–715
see also Mixe-Zoquean languages
 Sierra Totonac 1080
 nouns 1083
 speaker numbers 1081
see also Totonacan languages
 Sign language(s) 953–960
 acquisition 945
 children 945
 critical period hypothesis 945
 aspectual marking 952
 brain functions 945
 canonical forms 942*f*
 character signs 957–958
 Chinese finger/thumb negation 957–958
 classifiers 943, 943*f*, 946, 952
 signs 943, 943*f*, 952
 compounding 949, 949*f*
 current state of knowledge 953
 ages of languages 953–954
 basic vocabulary compilation 954
 links to other languages 954
 numbers of languages 953
 development 939–940
 difference degree, spoken language 952
 facial expressions 944, 944*f*, 958
 grammatical comparisons 958
 WH questions 944–945, 945*f*
 future developments 958
 grammatical similarities and differences 957
 headshake negation 958
 grammatical comparisons 958
 iconicity 946, 951–952
 modality 946, 946*f*
 morphology 951–952
 ideophones 946
 inflection 950, 951*f*
 inheritance 940
 iterative 941
 linguistic structure 940
 modality 939, 946
 visual-gestural modality 946, 946*f*
 monomorphemic signs 949, 949*f*
 morphology 949–953
 mouthing 958
 grammatical comparisons 958
 movement modification 950, 950*f*, 951
 negative affixation 949, 950*f*
 nominal verb derivation 950, 950*f*
 number incorporation 949–950, 950*f*
 phonology 941
 handshape organization 941, 941*f*
 minimal pairs 941, 941*f*
 plural sweep 951, 952*f*
 polymorphemic signs 952, 952*f*
 question words 958
 referential loci 942, 943*f*
 relationships between sign languages 956
 American Sign Language 956
 British Sign Language family 956
 colonial history, effect of 956–957
 creolization 956
 educational facility establishment 956–957
 family trees 957
 Japanese Sign Language family 956
 language mixing 956
 Old French Sign Language 956
 simultaneous morphology 957
 sociocultural and sociolinguistic variables 954
 continuous emergence 954–956
 foreign sign languages 954
 urban sign languages 954
 village-based sign languages 954
 spatial mechanisms 957
 syntax 944
 recursion 944
 WH questions 944
 verbal agreement 942, 943*f*
 verbal modification 951
 Sika
 classification 420
 voice alteration 420
see also Flores languages
 Sikhism, Punjabi 886
 SIL (Summer Institute of Linguistics)
 Arremte study 73
 Chico language studies 226
 Simultaneous morphology, sign languages 957
 Sinasina 1086–1087
see also Chimbu-Wahgi languages
 Sindhi 960–964
 cases 962
 classification 251–252, 522
 as Dravidian language 960–961
 dialects 961, 962
 future 963*t*
 gender 962, 963*t*
 grammars (books) 963
 habitual 963*t*
 history 960
 earliest reference 961
 influences from other languages 961
 morphology 962
 nouns 962
 phonology 961
 consonants 961, 962*t*
 stops 961
 syllable structure 962
 vowels 961–962
 postposition 962
 Qu'ran translation 961
 related languages 961
 Kachchi 961
 Siraiki 961
 SOV 962
 syntax 963
 use of 960
 number of speakers 523
 verbs 962, 963, 963*t*
 word order 963
 writing systems 524
see also Dardic languages; Gujarati; Indo-Aryan
 languages; Kashmiri
 Sindhu *see* Sindhi
 Singapore
 Austronesian languages 97
 Hindi 495
 Malay 679
 Mandarin Chinese 679
 official languages 679
 Tamil 679
 Singapore English 360, 679
 register 361
 Singhalese *see* Sinhala
 Singular nouns, in introflecting language 51
 Sinhala 964–968
 cases 965
 classification 251–252, 522
 clef/focused sentence construction 967
 conjunctive participles 966
 dative subject sentences 967
 demonstratives 965
 Dhivehi *vs.* 285
 genders 965, 966*t*
 influence from other languages 964

- Sinhala (*continued*)
 morphology 965
 nonverbal sentences 966–967
 nouns 966*t*
 orthography 965, 966*t*
 phonology 965
 consonants 965, 965*t*
 vowels 965, 965*t*
 postpositions 965–966
 pronouns 965
 script 964
 subject case forms 967
 syntax 965
 use of
 Buddhist traditions 964
 as official language 964
 Sri Lanka 964
 varieties 964
 spoken *vs.* literary 964–965, 966*t*
 word order 965–966
see also Dardic languages; Dhivehi; Indo-Aryan languages
 Sinhalese *see* Sinhala
- Sinitic languages
 classification 969
 subgroupings 1010*t*
 syllable structure 969–970
see also Sino-Tibetan languages
- Sino-Tibetan languages 968–971
 classification 253
 influence from other languages 970
 influence on other languages 970
 Karen languages 581
 subgroupings 968–969
 use of 968
see also Austric hypothesis;
 Austroasiatic languages; Burmese;
 Chinese; Karen languages;
 Proto-Sino-Tibetan; Sinitic languages;
 South Asian languages; Southeast
 Asian languages; Tibeto-Burman
 languages
- Siona
 accent/tone 1096
 adjectives 1099
 case markers 1097
 consonants 1095
 evidentiality 1100
 nasalization 1095
 plurals 1097
 speaker numbers 1092*t*
 syllable pattern 1095
 verbs, evidentiality 1100
see also Tucanoan languages
- Siouan languages 749
 argument structure 972
 demography 971
 dictionaries 973
 external relationships 972
 Catawban languages *vs.* 972
 lexicon 972
 locations 971
 morphology 972
 phonology 972
 postpositions 972
 SOV 972
 subgroups 971
see also Biloxi
- Sioux *see* Lakota
- Sipakapense 705–706
 speaker numbers 706*t*
see also Mayan languages
- Siraiki 635
 Sindhi *vs.* 961
- Siraya 421
see also Formosan languages
- Sireniki 373
 history 371
see also Eskimo-Aleut
- Siriano
 adjectives 1099
 animate noun classifiers 1097
 case markers 1096
 consonants 1094*t*
 evidentiality 1101
 future 1098*t*
 morphemes 1096
 nasalization 1095–1096
 plurals 1097
 speaker numbers 1092*t*
 verb evidentiality 1101
see also Tucanoan languages
- Siswati *see* Swati
- Siwai *see* Motuna (Siwai)
- Skene, W F 856
- Skolt Saami 911
see also Saami
- Skou languages 973–974
 classification 253, 973
 gender 974
 morphosyntax 973
 phonology 973
 consonants 973
 vowels 973
- SOV 258
 subject marking 974
 use of
 geographical distribution 840–841
 New Guinea 973
 verb agreement 974
 word order 973
see also Papuan languages;
 Warupu (Barupu)
- Slavic languages 974–977
 conjugation 977
 declension 976
 influence on other languages
 Esperanto 376
 Romanian 901
 morphology 976
 perfect 977
 phonology 975
 diphthongs 975
 prosody 976
 sonority 975
 syllabic synharmony 975
 vowels 976
 types 974–975
see also Balto-Slavic languages;
 Belorussian; Bulgarian; Czech;
 Indo-European languages;
 Macedonian; Polish; Russian; Slovak;
 Slovene; Sorbian; Turkic languages;
 Ukrainian
- Slavonic languages, Church *see* Church Slavonic
- Slovak 977–981
 adjectives 978–979
 biaspectual verbs 979
 cases 978
 classification 251–252
 consonant alternations 979
 declensions 976–977, 979
 dialects 980
 future 979
 genders 978
 imperfective 977, 979
 iterative 979
 lexicon 980
 word-borrowing 980
 morphology 978
 mutation 979
 nouns 978, 979
 origin/development 980
 orthography 977
 Latin alphabet 977–978
 phonology 978
 consonants 978
 diphthongs 978
 ‘rhythmic law,’ 978
 vowels 978
 word stress 978
 plurals 978–979
 pronouns 978–979
 resultative 975
 Stúr, L’udovit 980
 syntax 980
 use of 977
- verbs 979
 perfective verbs 979
 word order 980
 workers in, Bernolák, Anton 980
 written 980
see also Slavic languages; Slovene
- Slovakia
 German 444
 Hungarian 514
 Romani 898
- Slovene 981–985
 adjectives 983
 alphabet 982*t*
 cases 983
 classification 974–975
 as inflecting language 983
 clitics 984
 Croatian *vs.* 981
 declensions 976–977
 dialects 981, 982*f*
 genders 983
 grammars (books) 981
 imperfective 983
 lexicon 985
 maintenance 981
 morphology 983
 nouns 983, 984*t*
 origin/development 981
 phonology 982
 consonants 983, 983*t*
 stress patterns 983
 vowels 982, 983*t*
 word prosody 983
 writing systems 982
 political issues 981
 syntax 984
 use of 981
 Italy 545, 981
 as official language 981, 982
 Slovenia 981, 982
 verbs 983, 984, 984*t*
 word order 984
 writings 981
 Bible translations 981
 earliest documents 981
see also Balto-Slavic languages;
 Bulgarian; Macedonian; Slavic
 languages; Slovak
- Slovenia
 Hungarian 514
 Slovene 981, 982
- S6 (Tro) 726–727
see also Katuic languages
- Social dislocation 325
- Social distributions, Native American languages 746
- Social mobility, Later Modern English development 344
- Sociolect/social class, Arabic 55
- Sociolinguistics, Thai 1060
- Sogdian 537–538, 985–987
 alphabets 985
 classification 251–252
 declensions 540–541
 definite articles 540
 dual 540
 genders 540
 history 985
 imperfect tense 541
 modal forms 542
 morphology 986
 past tenses 541, 986
 perfect 986
 phonology 986
 ‘potentials,’ 986
 progressive 537–538
 SVO 984
 tenses 541
see also specific tenses
 use of 985
 verbs 986
 workers in, Pelliot, Paul 986
 written texts
 Buddhist texts 986

- differences 986
 oldest 985
 religious texts 985
see also Aramaic; Avestan; Iranian languages; Old Church Slavonic; Persian, Old; Tajik Persian
- Soghdian Inner Asian scripts *see* Manchu
- Solano 751
see also Native American languages
- Solresol 76
- Somali 987–990
 adjectives 988
 alienability 988
 complex sentences 989
 focus particle 989
 morphology 987
 nouns 988
 phonology 987
 phonemes 987, 987*t*
 syllable structure 987
 tone 987
 progressive 987
 questions 989
 sentence structure 988, 989
 SOV 989
 syntax 988
 use of 987
 number of speakers 272–273
 verbs 987
 word order 989
see also Afroasiatic languages; Cushitic languages; Ethiopian linguistic area (ELA)
- Somalia
 Cushitic languages 272–273
 Swahili 1026
- Somray 728
see also Pearic languages
- Song, Hopi 514
- Songai languages 990–991
 classification 773
 Greenberg, Joseph H 991
 as tonal languages 991
 development 991
 use of 990–991
 distribution 775*f*
 speaker numbers 772–773
 varieties 990–991
 vowel harmony 773–774
 word order 991
 workers in, Greenberg, Joseph H 991
see also Nilo-Saharan languages
- Songhay languages 253
see also Nilo-Saharan languages
- Sonoran 1139
see also Uto-Aztecan languages
- Sora (Savara) 736
see also Munda languages
- Sorbian 991–995
 alphabet 994
 classification 251–252, 974–975
 ‘dialect centers,’ 993
 grammar 994
 imperfective 994
 iterative 994
 perfect 994
 SOV 994
 ‘transitional dialects,’ 994
 use of 991–993, 992*f*
 current speakers 993
 Germany 991–993
 word order 994
 writings 993–994
see also Balto-Slavic languages; Czech; German; Polish; Slavic languages
- Sotho, Southern 1017–1018
 Lesotho 1017–1018
 South Africa 1017–1018
see also Sotho-Tswana languages
- Sotho-Tswana languages 1017
see also Bantu languages, Southern
- Souei (Proom) 726–727
see also Katuic languages
- Sougb
 classification 1176
 nominal complex 1177
see also East Bird’s Head (EBH) languages; West Papuan languages
- Sound change, and long-range comparison 650–651
- Sound correspondences
 and long-range comparison 650–651
 nongenuine 651
- Sound harmony
 Chuvash 244
 Turkic languages 1111
 Uzbek 1147
 Yakut 1200
- ‘Sound’ plurals, in introflecting language 52
- South Africa
 Afrikaans 7
 Fanagalo 411
 Gujarati 468
 language shift 321
 Ndebele 1187
 Northern Sotho 1017–1018
 Sepedi 1187
 Sesotho 1187
 Setswana 1187
 Siswati 1187
 South African Ndebele 1018
 Southern Bantu languages 1017
 Southern Sotho 1017–1018
 Swati 1018
 Tshivenda 1187
 Tsonga 1018
 Tswana 1017–1018
 Venda 1017
 Xhosa 1018
 Xitsonga 1187
 Zulu 1018
- South African Ndebele 1018
see also Nguni languages
- South America
 language families 651–652
 Native American languages *see* Native American languages
- South Asian Association of Regional Cooperation (SAARC) 522
- South Asian languages 62, 995–1001
 absolutive 62–63
 classifiers 997–998
 compound verbs 996
 converbs 995, 996–997, 998–999
 dative subjects 997
 historical evidence 999
 isoglosses 1000
 Kuiper, F B J 998–999
 Maisica, C P 999, 1000
 morphological causatives 997
 quotatives 997
 reduplication 998
 research history 998
 retroflex consonants 995
 Southworth, F C 999
 SOV 62–63
 subareas 1000
 word order 995
see also Austronesian languages; Balkan linguistic area; Bengali; Burushaski; Dravidian Languages; Europe, as Linguistic Area; Hindi; Indo-Aryan languages; Indo-European languages; Sino-Tibetan languages
- South Bird’s Head (SBH) languages
 classification 1176
 word order 1176
see also West Papuan languages
- South Bougainville languages 841
see also Papuan languages
- Southeast Asian languages 1009–1017
 classifiers 1013–1014, 1013*t*, 1014*t*
come to have verb 1015
 directional verbs 1014
 geography 1009
 grammaticalization 1012
 indeterminateness 1011
 Islam *see* Islam
- language families 1009
 nonobligatory categoricals 1011*t*
 pragmatics 1011
 structure 1012
 syllabic morphology 1011
 syntactic patterns 1014
 tense-aspect-modality (TAM) markers 1014
 versatility 1012
 word order 1013, 1013*t*, 1014, 1014*t*
see also Areal Linguistics; Austroasiatic languages; Balkan linguistic area; Europe, as Linguistic Area; Indo-Aryan languages; Kapampangan; Mon-Khmer languages; Sino-Tibetan languages; Tai languages
- Southern Altay Turkic 611
- Southern Arabic *see* Arabic, Southern
- Southern Bantu languages *see* Bantu languages, Southern
- Southern Sama *see* Sama, Southern
- Southern Sotho *see* Sotho, Southern
- Southern White Vernacular English (SWVE) 335
- South Halmahera/West New Guinea (SHWNG) languages 685
see also Malayo-Polynesian languages
- South Mindanao languages
 classification 1002*t*
 phonology
 consonants 1002
 vowel loss 1003
 vowels 1002
see also Malayo-Polynesian languages
- South Oghuz *see* Oghuz, South
- South Philippine languages 1001–1009
 antipassives 1004
 case-marking 1005
 classification 1001–1002, 1002*t*
 classifiers 1004
 clitics 1004
 ergative 1002
 grammar 1002
 morphemes 1003–1004
 morphology 1003
 morphosyntax 1004
 phonology 1002
 consonants 1002, 1003
 vowels 1002
 sentences 1004
 syntax 1008
 markers 1003–1004
 use of 1001–1002
 verbal affixes 1004
 word order 1004, 1004*f*
see also Austronesian languages; Bikol; Ilocano; Malayo-Polynesian languages; North Philippine languages; Tagalog
- South Saami *see* Saami, South
- Southwestern Mandarin
 classification 214
 speaker numbers 214*t*
see also Mandarin
- Southworth, F C 999
- SOV
 Ainu 15–16
 Amharic 36
 Australian languages 90
 Bashkir 143
 Basque 146
 Bengali 148
 Burushaski 178
 Creek 266
 Cushitic languages 275
 Eskimo-Aleut languages 371–372
 Ethiopian linguistic area (ELA) 380
 Evenki 406
 German 446
 Germanic languages 449
 Highland East Cushitic (HEC) languages 489
 Hindi 497
 Hokan languages 507
 Hopi 511
 Ijó 517–518
 Indo-Iranian languages 534
 Inupiaq 536
 Japanese 558

- SOV (*continued*)
 Kannada 577
 Khoasaan languages 602
 Luxembourgish 659–660
 Madang languages 671
 Malayalam 683
 Mande languages 697
 Marathi 704
 Middle English 354
 Munda languages 736–737
 Navajo 761
 Nenets (Yurak) 763
 Omaha-Ponca 804
 Oromo 812
 Ossetic 817
 Oto-Mangean languages 823–824
 Papuan languages 841
 Pashto 848
 Punjabi 889
 Saami 911
 Sindhi 962
 Siouan languages 972
 Skou languages 258
 Somali 989
 Sorbian 994
 South Asia 62–63
 Tai languages 1040
 Tanoan languages 1048–1049
 Telugu 1058
 Thai 1059
 Tibetan 1062
 Tigrinya 1065
 Toda 1072
 Trans New Guinea languages 1087
 Tungusic languages 1104
 Tupian languages 1107–1108
 Turkic languages 1111–1112
 Turkish 1115–1116
 Uralic languages 1132
 West Papuan languages 1176
 Wolaitta 1183
 Yukaghir 1211–1211
- Spain
 Basque 144–145
 Catalan 188
 Galician 435
 Spanish 1020
- Spanglish *see* Yanito
- Spanish 1020–1022
 borrowing 651–652, 653
 classification 251–252
 concord 1021
 diminutives 1021
 history 1020
 indicative mood 1021
 influence on other languages
 Keres 591
 Korean 615–616
 Krio 620
 Mapudungan languages 702
 Mayan languages 708–709
 Palenquero 828
 Tagalog 1036
 Tohono O'odham 1074
 Yanito 1202
 influences from other languages 1020
 Arawak languages 59
 Mayan languages 708–709
 morphology 1021
 nouns 1021
 OVS 884
 perfect 1021
 phonetics 1020
 consonants 1020–1021
 semi-vowels 1020–1021
 phonology 1020
 plurals 1021
 subject 1021
 subjunctive 1021
 syntax 1021
 use of 1020
 verbs 1021
 irregular verbs 1021
 vocabulary 1021
- word order 1021
see also Basque; Catalan; Indo-European languages; Latin; Portuguese; Romance languages; Yanito
- Spanyol *see* Judeo-Spanish
- Spanyolit *see* Judeo-Spanish
- Spatial mechanisms, sign languages 957
- Spatial orientation
 Nuristani languages 788
 Warlpiri 1167–1168, 1167*t*
- Speak Good English* movement 361–362
- Sprachbund
 definition 119
see also Linguistic areas
- Sprachbund, language diffusion 248
- Sranan
 classification 249*t*
 influence from other languages 858
 lexicon 858
- Sri Lanka
 Indo-Aryan languages 522
 official language 964
 Pāli 830
 Sinhala 964
- Standard Average European (SAE) languages
 392–393
 characteristics 393–394
- Statenbijbel* 308
- Stein, Aurel 1068–1069
- Steinthal, Heymann, Mande language
 classification 696
- Stem morphophonological alternations, in
 agglutinating languages *see* Finnish
- Stevens, Thomas 921
- Stieng 726
see also Bahnaric languages
- Stigmatization, nonnative English 361
- Stød, Danish pronunciation 280
- Strehlow, Carl, Arrernte study 73
- Strehlow, T G H 73
- Stress
 Achagua 60
 Arapaho 26
 Arawak languages 60
 Bakairi 183–184
 Balkan linguistic area 122
 Baure 60
 Breton 167
 Cariban languages 183–184, 185*t*
 Caucasian languages 194
 Cayuga languages 543
 Cheyenne 26
 Chinantec 212
 Cree 26
 Dutch 309
 English, Modern 328
 Finnish (Suomi) 414
 Hungarian, phonology 515
 Ilocano 519
 Iroquoian languages 543
 Italian 551
 Kaytetye 587
 Korean 614
 Kuikuro 183–184
 Latvian 645
 Lithuanian 646
 Montagnais 26
 Ojibwa 26
 Panare 183–184
 Polish, phonology 875
 Proto-Algonquian 26
 Punjabi 887
 Romanian 901–902
 Saami 912
 Slovak 978
 Slovene 983
 Tariana 60
 Thai 1059
 Tohono O'odham 1075
 Tupian languages 1106
 Turkish 1113
 Uralic languages 1131
 Wakashan languages 1158
 Warekena (Guarequena) 60
- Waurá 60
- Yiddish 1204
- Yukpa 183–184
- Stress accents, Greek, Modern 466
- Stressed schwa, Balkan linguistic area 122
- Strict agreement markers, Standard Average
 European (SAE) languages 393–394
- Stúr, L'udovit, Slovak 980
- Style, North American native language variation
 758
- Subanon languages
 classification 1001–1002, 1002*t*
 consonants 1003
 syntax markers 1003–1004
see also South Philippine languages
- Subareas, South Asian languages 1000
- Subgroups, genetic classification 246
- Subjunctives, analytic, Balkan linguistic area 127,
 128*t*
- Substrate theory, Creole origins 859
- Subtiaba-Tlapanec languages 751
see also Oto-Mangean languages
- Sudan
 Adamawa-Ubangi languages 771
 Dinka 293
 Fulfulde 430
 Kordofanian 770
 Kordofanian languages 613
 Nilo-Saharan languages 774
- Sudre, Francois 76
- Suffix(es)
 diachronic origins 287
 in isolating language 222, 222*t*
 morphophonological alternations, in
 agglutinating languages *see* Finnish
 preference 288
 prefixation *vs.* 288
 Uzbek 1147
 workers in 288
- Sukhothai dialect, Thai 1059
- Sulawesi
 Austronesian languages 99
 Javanese 560
- Sulka 841
see also Papuan languages
- Sum 282
see also Pashai languages
- Sumatra, Javanese 560
- Sumbawa languages, Austronesian
 languages 99
- Sumerian 929, 1022–1026
 classification 249
 clauses 1024
 earliest sources 1022
 habitual 1024
 morphology 1023
 noun phrases 1023
 case markers 1023
 genitive cases 1023
 prefixes 1023
 nouns 1023
 compounding 1023
 gender 1023
 phonology 1022
 consonants 1022–1023
 vowels 1022–1023
 possessives 1024
 resources 1025
 use of 1022
 verbs 1024
 adverbs 1024–1025
 aspect categories 1024
 clitics 1025
 finite verbs 1025
 irregular 1024
 stative verbs 1024
 tense categories 1024
 word classes 1023
 determiners 1023
see also Akkadian; Babylonian; Eblaite;
 Elamite; Semitic languages
- Summer Institute of Linguistics *see* SIL (Summer
 Institute of Linguistics)
- Sumo Tawahka *see* Sumu (Sumo Tawahka)

- Sumu (Sumo Tawahka)
 classification 711
 dialects 711
 use of 711
see also Misumalpan languages
- Sunda *see* Sundanese (Sunda)
- Sundanese (Sunda) 99
see also Austronesian languages
- Suomi *see* Finnish (Suomi)
- Suoy 728
see also Pearic languages
- Superlatives, in introflecting language 52
- Superstrate theory, Creole origins 860
- Suriname
 Arawak languages 59
 Dutch 307
 Javanese 560
- Surmic
 classification 773
 use of 775*f*
see also Nilo-Saharan languages
- Surui 1106*t*
see also Tupian languages
- Susu 620
- Sutta Pitaka*, Pāli canonical texts 831–832
- Svan, dialects 193
- Sverdrup, Harald V 856
- SVO
 Adamawa-Ubangi languages 3
 African languages 5
 Afrikaans 9
 Akan 19
 Arabic 47
 Balkan linguistic area 131
 Bantu, Southern 1019
 Bantu languages 141–142
 Basque 146
 Benue-Congo languages 151
 Coptic 39
 Dinka 294
 English 341
 Esperanto 376
 Fanakalo 412
 Finnish (Suomi) 413
 French 429
 Gikuyu (Kikuyu) 451
 Gur languages 473
 Hausa 478
 Hawaiian Creole English (HCE) 481
 Italian 554
 Karen languages 581
 Kashmiri 583–584
 Khasi languages 595–596
 Khotanese 604
 Kinyarwanda 607
 Kru languages 624
 Kwa languages 632
 Lao 639–640
 Luo 659
 Malukan languages 690
 Mambila 692
 Mon 720
 Munda languages 737
 Norwegian 785
 Papiamentu 835
 Papuan languages 841
 pidgins 862
 Portuguese 884
 Russenorsk 904
 Sogdian 984
 Thai 1059
 Tok Pisin 1077
 Torricelli languages 1078
 Tungusic languages 1104
 Turkish 1115–1116
 Western Songai 991
 West Papuan languages 1177–1178
 Zulu 1215–1216
- Swadesh, Morris, Oto-Mangean language
 classification 819
- Swahili 1026–1030
 agreement 1027
 classifications 137, 253
 concord 1027–1028
 dialects 1026–1027
 diminutives 1027
 as first language 1026
 Gikuyu, influence on 449–450
 history 1026
 location inversion structures 1029
 as national language 1026
 noun classes 1027
 noun phrase 1028–1029
 nouns 1027
 possessives 1027*t*
 subject markers 1027–1028
 suffixes 1028
 syntax 1028
 use of 1026
 verbs 1027–1028
 word order 1028–1029
see also Bantu languages; Gujarati; Niger-Congo languages
- Swati 1017
 South Africa 1018
 Swaziland 1018
 Zulu *vs.* 1215
see also Nguni languages
- Swat-Kohistani, phonology, sibilants 283
- Swaziland
 official languages 1018
 Southern Bantu languages 1017
 Swati 1018
- Sweden
 Estonian 377
 Finnish 413
 Saami 911
 Swedish 1030
 Urdu 1133
- Swedish 1030–1033
 classification 251–252
 dialects 1032
 new varieties 1032
 noun phrases 1031
 orthography
 alphabet 1030
 runes 1030
 participles 1031
 perfect 1030
 phonology 1030
 consonants 1030
 tonality 1030
 vowels 1030
 possessives 1031
 subordinate clauses 1032
 use of 1030
 as verb-L2 1032
 verbs 1030
 word order 1032
see also Germanic languages; Icelandic; Norse; Old; Norwegian; Scandinavian languages
- Sweet, Henry, Later Modern English definition 343
- Switzerland
 French 427
 German 444
 Italian 545
 Romansh 893
- Syllable(s)
 morphology 1011
 Thai 1059
- Syntactic patterns, Southeast Asian languages 1014
- Syria
 Domari 295
 Kurdish 538, 625
- Syriac 58, 1033–1034
 classification 250
 lexicon 1034
 morphology 1033
 nouns 1033
 origin/development 1033
 perfect 1033
 phonology 1033
 consonants 1033
 vowels 1033
 pronouns 1033
 religious uses 1033
 root-and-pattern 1033
 sentence structure 1034
 use of 1033
 verbs 1033
 writing system 1033
see also Afroasiatic languages; Arabic; Aramaic; Hebrew; Modern Standard Arabic (MSA); Semitic Languages; Semitic languages
- Syriac Christianity 1033
see also Aramaic
- Syrian Orthodox Church, languages, Aramaic 58
- ## T
- Taalmonument, Afrikaans 9, 9*f*
- Tadzhik *see* Tajik Persian
- Tagalog 1035–1038
 Cebuano *vs.* 197
 consonants 1036
 derivational affixes 1037
 diphthongs 1036
 glottal stops 1036
 grammar 1036
 growth of use 1035–1036
 influence on other languages 1037, 1038
 iterative 1036–1037
 loanwords, Spanish 1036
 as official language 1035
 origin/development 1035
 phonology 1036, 1036*t*
 Samar-Leyte *vs.* 916
 spelling 1036
 as synthetic language 1036
 use of 99, 1035
 Philippines 783, 1035
 verbs 1036
 tenses 1036–1037
 vowels 1036
see also Austronesian languages; Cebuano; Hiligaynon; Kapampangan; North Philippine languages; Samar-Leyte; South Philippine languages
- Tagoy 613
see also Kordofanian languages
- Tahitian 1038–1039
 classification 250–251
 as official language, French Polynesia 1039
 phonemes 1039
 use of 1038–1039
see also Hawaiian; Oceanic languages
- Tai Daeng *see* Red Tai (Tai Daeng)
- Tai-Kadai (Zhuang-Dong) languages
 classification 968
 lexicostatistics 248
 use of 105
see also Sino-Tibetan languages
- Tai languages 1039–1041
 affiliations 1039
 Tai-Kadai link 1039–1040
 classification 1039
 history 1040
 loan words 1040
 SOV 1040
 subgroupings 1010*t*
 as tonal languages 1040
 types 1040
 use of 1039
 VSO 1039
 word order 1040
 writing system 1040
see also Austro-Tai hypothesis; Southeast Asian languages; Thai
- Taiwan
 Austronesian languages 97, 105
 Cebuano 197
 Sino-Tibetan languages 968
- Taiwanese
 classification 969
see also Hakka languages
- Tajik *see* Tajik Persian
- Tajiki *see* Tajik Persian

- Tajikistan
 Indo-Iranian languages 531
 Kazakh 588
 Kirghiz 610
 Modern Persian 538, 850
 Tajik Persian 1041
 Uzbek 1145
- Tajik Persian 1041–1044
 classification 251–252
 classifiers 1042
 future 1043
 gender 1042
 history 1041
 written language 1041
 influence from other languages, Russian 1043
 lexicon 1043
 causatives 1043
 conjunct verbs 1043
 denominal verbs 1043
 prefixes 1043
 suffixes 1043
 morphology 1042
 noun phrase syntax 1042
 orthography 1041
 Cyrillic alphabet 1041–1042
 vowels 1041–1042
 perfect 1042
 personal pronouns 1042
 phonology 1041
 consonants 1041
 Uzbek *vs.* 1041
 vowels 1041*f*, 1041–1042
 postpositions 1042
 progressive 1042
 syntax 1043
 use of 1041
 Uzbekistan 1041, 1145
 verbs 1042
see also Iranian languages; Persian, Modern;
 Persian, Old; Russian; Turkic languages;
 Uzbek
- Takelma 653
 Takelma-Kalapuya 750
see also Penutian languages
- Takhaht *see* Nuuchahnulth
- Takic languages 1139*t*
see also Uto-Aztecan languages
- Takpa 968–969
see also Bodish languages
- Talassa 613
see also Kadugli languages
- Talla 613
see also Kadugli languages
- Tallán-Sechura 41
- Talmud 483
- Talodi 770
see also Kordofanian
- Tamambo 1044–1047
 affixation 1046
 classifiers 1047
 compounding 1046
 as first language 1044
 grammar 1044
 individuation 1047
 lexicon 1046
 orthography 1046
 phonology 1046
 possessive constructions 1047
 reduplication 1046
 serial verb constructions 1046
 use of 1044
 valency changing affixes 1046
 word order 1046
see also Austronesian languages; Language
 endangerment
- Taman 775*f*
- Tamanaku 185*f*
see also Cariban languages
- Tamang 968–969
see also Bodish languages
- Tamil 1047–1049
 agreement 299*t*
 classification 251
 consonants 298
 converbs 995
 dative subjects 997
 grammar 1048–1049
 influence on other languages, Malayalam
 680–681
 Malayalam *vs.* 682
 nouns
 ablative 302
 accusative 301–302
 genitive 302
 nominative case 301
 origin/development 1047–1048
 personal suffixes 305*t*
 phonology 1048
 postpositions 1048–1049
 pronouns 302–303, 303*t*
 religious influences 1047–1048
 script *see* Tamil script
 tenses 304
 use of, Singapore 679
see also Dravidian languages; Malayalam
- Tamil script 297, 1048
 earliest examples 1047–1048
see also Malayalam
- Tangik languages 250
see also Australian languages
- Tangsa (Naga) 968–969
see also Konyak languages
- Tani languages 968–969
see also Adi; Apatani; Sino-Tibetan languages
- Tanimuca *see* Retuara/Tanimuca
- Tanoan languages 750
 external relationships 1049
 future work 1050
 grammatical features 1049
 historical aspects 1049
 locations 1049
 nouns 1049–1050
 phonology 1049
 four-way stop contrast 1049–1050
 SOV 1048–1049
 speakers 1049
 subgroups 1049
 Uto-Aztecan language link 1140
 verbs 1049–1050
 word order 1049–1050
see also Hope-Tewa
- Tano languages 631
 verbs 632
 vowel harmony 632
see also Abure; Ahanta; Anufo; Anyi; Kwa
 languages
- Tanzania
 Cushitic languages 272–273
 Luo 658
 national languages 1026
 Swahili 1026
- Ta'oih 726–727
see also Katuic languages
- Taokas-Babuza 421
see also Formosan languages
- Taracahitan 1140
see also Uto-Aztecan languages
- Tarascan 748
see also Native American languages
- Targum *see* Judeo-Aramaic
- Targumim, Jewish Palestinian Aramaic 58
- Tariana 1050–1052
 adjectives 1051
 causatives 1051
 classification 252–253
 classifiers 61, 1051
 evidentiality 1051
 genders 61, 1051
 instrumental case 1051
 locative case 1051
 morphology 1051
 nouns 1051
 origin/development 1050
 phonology 1050–1051
 plurals 1051
 as polysynthetic language 1050
 predicate structure 60
 pronominal suffix loss 60
 serial verb constructions 1051
 stress 60
 switch-reference 1051
 tenses 1051
 Tucanoan languages *vs.* 1051
 use of 1050
 verbs 60, 1051
see also Arawak languages
- Tasmania 86
- Tatar 1052–1055
 contacts 1052–1053
 Chaghatay 1053
 Kuman 1053
 Ottoman 1053
 converbs 1054
 dialects 1054
 distinctive features 1053
 grammar 1054
 history 1052
 lexicon 1054
 location 1052
 origin 1052
 phonology 1053
 consonants 1053–1054
 vowels 1053
 possessives 1054
 related languages 1053
 speakers 1052
 use of 1052
 written language 1053
see also Bashkir; Turkic languages
- Tatarstan, languages 1052
- Tataviam 1140
see also Uto-Aztecan languages
- Tatuyo
 accent/tone 1096
 case markers 1096
 consonants 1094
 personal pronouns 1098
 speaker numbers 1092*t*
 verbs 1099–1100
 compound verb roots 1100
see also Tucanoan languages
- Taulil 841
see also East New Britain languages
- Tavgy *see* Nganasan (Tavgy)
- Tboli
 case marking 1006–1007, 1007*t*
 morphosyntax 1006
 negation 1007*t*
 phonology, vowel loss 1003
see also South Mindinao languages
- Tebriz 112–113
see also Azerbaijani
- Tectiteco *see* Teko (Tectiteco)
- Tedim *see* Tiddim (Chin: Tedim)
- Tedim (Tiddim: Chin) 968–969
see also Kuki-Chin languages
- Tegali 613
see also Kordofanian languages
- Tegem 613
see also Kordofanian languages
- Teko (Tectiteco) 705–706
 speaker numbers 706*t*
see also Mayan languages
- Tektiteco *see* Teko (Tectiteco)
- Telugu 1055–1058
 adjectives 1057
 adverbs 1057
 agreement 299, 300*t*
 classification 251
 concord 1055–1056
 consonants 1055*t*
 influence from other languages 1055, 1058
 nouns 1056
 case 1056
 genitive 302
 instrumental case 302
 number 1056
 plural suffixes 301
 numerals 303, 1056
 oblique forms 1056, 1056*t*
 personal suffixes 305*t*
 phonology 1055

- postpositions 1056
 pro-drop 1058
 pronouns 302–303, 303*t*, 1055
 honorifics 1056, 1056*t*
 script *see* Telugu script
 SOV 1058
 syntax 1058
 use of, India 1055
 verbs 304, 1057
 compounding 1058
 conjugation classes 1058
 inflected verbs 1057
 nonfinite verbs 1057–1058
 pronominal suffixes 1057
 tense/mood 1057
 vocabulary 1058
 vowel harmony 1055
 vowels 1056*t*
 vowel harmony 1055
 word order 1058
 writing *see* Telugu script
see also Brahui; Dravidian languages;
 Malayalam
 Telugu script 297
 Temein 773–774
see also Nilo-Saharan languages
 Temne 770
 Krio, influences on 618
see also Atlantic Congo languages
 Temporal forms, Nuristani languages 787
 Tense and aspect
 Arabic 47
 Creoles 862
 Cushitic languages 274–275, 275*f*
 Evenki 406
 Hausa 478
 Indo-Iranian languages 533
 Kinyarwanda 606, 608
 Pitjantjatjara 872–873
 Southeast Asia 1014
 Tense-aspect-modality (TAM) markers, Southeast
 Asian languages 1014
 Tense markers, language diffusion 248
 Tepehua *see* Totonacan languages
 Tepiman 1140
see also Uto-Aztecan languages
 Tepo-Plapo 624
see also Grebo languages
 Tequistlatecan 751
see also Hokan languages
 Tequistlateco 748
see also Native American languages
 Terëna 60
see also Arawak languages
 Ter Saami 911
see also Saami
 Te'utujil 709*t*
see also Mayan languages
 Tewa 1049
 phonology 1049–1050
 Texistepec 714
see also Mixe-Zoquean languages
 'Thaana,' Dhivehi 285
 Thai 1058–1060
 classification 253
 classifiers 1060
 distribution 1058–1059
 future work 1060
 historical aspects 1059
 Sukhothai dialect 1059
 influence on other languages
 Khmer (Cambodian) 600
 Lao 640
 loanwords 1059
 Khmer 1059
 Pali 1059
 Sanskrit 1059
 as national language 1039
 national language, Thailand 1058
 noun phrase 1059
 OSV 1059
 Pali, influences from 248
 particles 1060
 phonology 1059
 stress 1059
 syllables 1059
 tones 1059
 vowels 1059
 regional dialects 1058
 sociolinguistics 1060
 SOV 1059
 SVO 1059
 syntax 1059
 verbal predicates 1059–1060
 word order 1014*t*
see also Tai-Kadai (Zhuang-Dong) languages;
 Tai languages
 Thailand
 Aslian languages 94–95
 Burmese 170
 Karen languages 581
 Khmer (Cambodian) 597
 Khmuic languages 727
 Lao 639
 Malay 679
 Mon 718, 727
 Mon-Khmer languages 725
 national languages 1039
 Nyahkur 727
 Palaung-Wa languages 727
 Pali 830
 Sino-Tibetan languages 968
 Tai-Kadai 105
 Tai languages 1039
 Thai 1058
 Thamudic 932
see also Semitic languages
 Thao
 classification 421
 research history 423
see also Formosan languages
 Tharrkari 570
 Thavung-Phon Sung languages 728–729
see also Viet-Muong languages
 Theravada, languages *see* Pāli
 Thiin 570
 Tho (Tây) 1039
see also Tai languages
 Thomann, Georges 623
 Three-letter language identifiers 386
 Grimes, Joseph E 386
 International Organization for Standardization
 (ISO) 386
 Tibetan 1060–1063
 classification 968–969
 clauses 1062
 concord 1061
 dialects 1061
 future 1061
 grammar 1061
 honorifics 1062
 influence from other languages 1062–1063
 lexical verbs 1061
 noun phrases 1061
 past-tense clauses 1062
 phonology 1062
 central dialects 1062
 southern dialects 1062
 western dialects 1062
 present-tense clauses 1062
 recent history 1062
 sample sentence 1062
 SOV 1062
 tenses 1061–1062
 use of 1060–1061
 verbs
 verb phrases 1061
 verbs of being 1061
 vowel harmony 1062
 word order 1062
 words 1061
see also Bodish languages
 Tibeto-Burman languages
 classification 253
 Karen languages 581
see also Sino-Tibetan languages
 Tiddim (Chin: Tedim) 968–969
see also Kuki-Chin languages
 Tigré 929
 use of 382–383
see also Ethiopian Semitic languages; Semitic
 languages
 Tigrinya 1063–1065
 converbs 1064–1065
 morphology 1063
 nouns 1063–1064
 verbs 1064
 phonology 1063, 1064*t*
 SOV 1065
 syntax 1065
 use of 382–383, 929, 1063
see also Afroasiatic languages; Ethiopian
 linguistic area (ELA); Ethiopian Semitic
 languages; Semitic languages
 Tillamook 749
see also Salishan languages
 Timor-Altar-Pantar (TAP) languages
 classification 1087, 1176
 verbal complex, word order 1176
see also Trans New Guinea languages
 Timote-Cuica *see* Arawak languages
 Timucua 749
see also Muskogean languages
 Tindale, Norman 438
Tipitaka 831–832
 Tirahi
 classification 282
 sibilants 283
see also Kohistani languages
 Tiriyo
 geographical distribution 185*f*
 reduplication 184
 vowels 183–184
see also Cariban languages
 Tiro 613
see also Kordofanian languages
Tirukkural 1047–1048
 Tiv 253
see also Benue-Congo languages
 Tiwa 1049
 phonology 1049–1050
 Tiwi 1065–1068
 classification 250
 history 1065
 language changes 1065
 Modern Tiwi 1065–1066, 1067
 morphology/syntax 90
 New Tiwi 1065–1066, 1067
 isolating verbs 1067
 nouns 1067
 phonology 1067
 pronouns 1067
 vocabulary 1067
 word order 1067
 Traditional Tiwi 1065–1066
 adjectives 1066–1067
 consonants 1066, 1066*t*
 nouns 1066–1067
 plurals 1066–1067
 as polysynthetic language 1066–1067
 verb phrase 1066–1067
 verbs 1066–1067
 vowels 1066
see also Australian languages; Central Siberian
 Yupik; Creoles; Pidgins; Polysynthetic
 languages
 Tlachichilco Tepehua 1081
 speaker numbers 1081–1082
see also Totonacan languages
 Tlapanean, time depth 819
 Tlapaneko-Mangean languages 819–821
see also Oto-Mangean languages
 Tlapaneko-Sutiaba languages 819–821
see also Oto-Mangean languages
 Tlingit 252
see also Na-Dene languages
 Tocharian 1068–1071
 Buddhism 1069
 Celtic *vs.* 1070
 classification, genetic classification 246
 genders 1069–1070
 Germanic languages *vs.* 1070

- Tocharian (*continued*)
 Indo-European languages *vs.* 1070
 influences from other languages 1070
 manuscripts 1068–1069
 morphology 1069
 nominal compounds 1070
 nouns 1069–1070
 numerals 1070
 phonology 1069
 Proto-Indo-European languages *vs.* 1069
 reconstruction 246
 Tocharian A (Osttocharisch) 1068
 Tocharian B (Westtocharisch) 1068
 verbs 1070
 workers in 1068–1069
 writing system 1069
see also Chinese; Indo-Aryan languages; Indo-European languages; Iranian languages; Sanskrit; Turkic languages
- Tocho 613
see also Kordofanian languages
- Toda 1071–1074
 classification 251
 concord 1072
 consonants 298
 definition 1071
 as endangered language 1071
 Malayalam *vs.* 682
 modifiers 1072
 nouns 1072
 plural suffixes 301
 numerals 1072
 oblique forms 1073
 personal suffixes 305
 phonology 1071
 consonants 1071–1072, 1072*t*
 phonemes 1071
 vowels 1071–1072, 1072*t*
 pronouns 1072
 sentences 1072
 SOV 1072
 verbs 1073
 auxiliary verbs 1073
 bases 1073
 morphophonemic alternants 1073
 suffixes 1073
 tenses/modes 1073
 vocabulary 1073
 loanwords 1072
see also Dravidian languages
- Togo
 Ewe 408
 Gur 770
 Gur languages 472
 Kwa languages 771
 Mande 769–770
 Yoruba 1207
- Togo Mountain languages 631
 concord 632
 nouns 632
 subgroups 631
 vowel harmony 632
see also Adele; Animere; Kwa languages
- Tohono O'odham 1074–1076
 classification 252
 dialects 1074
 ergative 1075
 future 1075
 imperfective 1075
 influence from other languages 1074
 kin terms 1075
 morphology 1075
 nouns 1075
 orthography 1074
 phonology 1075
 consonants 1075
 stress 1075
 vowels 1075
 possessives 1075
 research 1074
 syntax 1075
 use of 1074
 Mexico 1074
 USA 1074
- verbs 1075
 VSO 1075
 word order 1075
see also Uto-Aztecan languages
- Tojiki *see* Tajik Persian
- Tojolab'al 705–706
 speaker numbers 707*t*
see also Mayan languages
- Tok Pisin 1076–1078
 classification 249*t*
 future 1077
 influence from other languages 1077
 German 858–859
 lexemes 1077
 lexicon 858–859
 linguistic relations 1076
 as national language 1076–1077
 origin/development 1076
 phonology 1077
 SVO 1077
 use of 1076–1077
 Papua New Guinea 836, 1076–1077
see also Creoles; Manambu; Pidgins
- Tolai 1077
- Tolkaappiyam 1047–1048
- Tolkien, J R R 76
- Tol languages 504
 classification 506
see also Hokan languages
- Tolubi 613
see also Kadugli languages
- Tonal languages, Chadic languages 206
- Tone
 Abun 1177
 Akan 19, 632
 Arawak languages 60
 Arikém 1106
 Bantu languages 139–140
 Burmese 172
 Carapana 1096
 Cherokee 543
 Chinantec 213*t*
 Chinese 215, 217*t*
 Chorotegan 821
 Danish 280
 Desano 1096
 Dinka 294
 Dogon 294
 Efik *see* Efik
 Ewe 408
 Ga-Dangme 632
 Gikuyu (Kikuyu) 450
 Gur (Voltaic) languages 473
 Hausa 478
 Iroquoian languages 543
 Japanese 558
 Jurúna 1106
 Kanuri 578
 Ket 593
 Kinyarwanda *see* Kinyarwanda
 Koyra Chiini 774
 Kpelle 697–698
 Krio 622
 Kru languages 624
 Kwa languages 632
 Luganda 657
 Macuna 1096
 Mambila 692
 Mandarin Chinese 223
 Mande languages 697
 Mano 697–698
 Matbat 1177
 Ma'ya 1177
 Meyah 1177
 Mohawk 543
 Mondé 1106
 Mpur 1177
 Mundurukú 1106
 Nilo-Saharan languages 774
 Oromo 810
 Oto-Mangean languages 821
 Putonghua 215, 217*t*
 Ramaráma 1106
 Resigaro 60
- Retuara/Tanimuca 1096
- Secoya 1096
- Sembla 697–698
- Shona languages 938
- Siona 1096
- Somali 987
- Tatuyo 1096
- Teréna 60
- Tucanoan languages 1096
- Tuparí 1106
- Tupian languages 1106
- Tuyuca 1096
- Vietnamese 1150, 1150*t*
- Waimaja/Bará 1096
- West Papuan languages 1177
- Wolaitta 1180
- Xipáya 1106
- Yoruba 1208
- Yuruti 1096
see also Obligatory Contour Principle (OCP)
- Tone-accent languages 810
- Tones
 Karen languages 581
 Thai 1059
- Tonga, Austronesian languages 102
- Tongan 250–251
see also Oceanic languages
- Tonkawa 751
see also Native American languages
- Tononacan *see* Native American languages
- Topic
 Chinese, spoken 216
 English, nonnative 361
 Japanese 558
- The Torah, Judaism 483–484
- Torkmancay 112–113
see also Azerbaijanian
- Torres Strait Islander 79
- Torricelli languages 1078–1080
 classification 253
 class systems 1078
 concord 1078
 diversity within 1078
 history 1079
 nominal inflexions 842
 noun classes 842
 phonetics 1078
 vowels 1078
 plurals 1078
 SVO 1078
 use of 1078
 geographical distribution 840
 Papua New Guinea 1078
- verbs
 morphology 1078
 serial verbs 1079
 voice system 1078–1079
 word order 1078, 1176
see also Arapesh (Bukiyip: Muhiang); Papuan languages
- Torwali
 agreement patterns 284
 sibilants 283
 speaker numbers 282
- Totoguañ 1074
see also Tohono O'odham
- Totonac 653
- Totonacan languages 748
 applicative affixes 1083–1084
 body part prefixes 1083
 classification 252
 imperfective 1083
 inflectional affixes 1083
 morphology 1083
 nouns 1083
 numerals 1083
 object agreement 1084
 phonology 1082
 consonants 1082*f*
 vowels 1082
 relationships 1081*f*
 syntax 1083
 Tepehua 1080

- Totonac 1080
 use of 1080f
 Mexico 1080
 verbs
 reciprocal verbs 1083
 verbal derivation 1083
 verbal inflexion 1083
 VSO 1084
 word order 1084
see also Native American languages
- Totontepec
 dependent verb forms 715
 phonology 713
 unstressed vowel loss 713
see also Mixe-Zoquean languages
- Totoró *see* Barbacoan languages
- Touo (Baniata) 841
 classification 204
 gender 205
 numbers 205
 phonology 205
 use of 204
see also Central Solomons languages
- Towa 1049
 phonology 1049–1050
see also Tanoan
- Traditional Tiwi *see* Tiwi
- Trager, George, Uto-Aztecan languages 1140
- Trans New Guinea languages 840
 classification 253, 1176
 concord 842
 conjunctions 1089
 dictionaries 1085–1086
 diversity 843
 grammar 1087
 grammars (books) 1085–1086
 hypothesis 1086
 cognate sets 1086, 1086t
 inflected verbs 1089
 medial verbs 842
 numbers 1089
 OVS 1087
 phonology 1087
 nasals 1087
 vowels 1087
 predicates 1089
 pronouns 1087–1089
 semantics 1087
 SOV 1087
 subgroups 1086
 suffixes 1089
 use of 1085
 geographical distribution 1088f
 speaker numbers 1085
 verb root 1087
 word order 1087
see also Angan languages; Asmat-Kamoro languages; Austronesian languages; Awyu-Dumut languages; Madang languages; Papuan languages; Proto Trans New Guinea language
- Transport, Later Modern English
 development 344
- Tree diagrams, genetic classification 246
- Tréma, French orthography 428
- Triki languages
 classification 819–821
 syllable onsets 821–822
see also Oto-Mangean languages
- Trinidad and Tobago, Hindi 495
- Trique 751
see also Mixtecan languages
- Tsachila *see* Barbacoan languages
- Tsakanian dialect 465
- Tsamosan 749
see also Salishan languages
- Ts'e-heng (Dioi) 1039
see also Tai languages
- Tshangla (Sharchopkha) 968–969
see also Bodish languages
- Tshivenda *see* Venda
- Tshwa 1018
 Mozambique 1018
 speaker numbers 1018
- Zimbabwe 1018
see also Ronga
- Tsimshian 750
see also Penutian languages
- Tsonga 1018
 Mozambique 1018
 South Africa 1018
 speaker numbers 1018
see also Ronga
- Tsotsi Taal 1090–1091
 definitions 1090
 development 1090
- Tsou
 classification 421
 research history 422–423
see also Formosan languages
- Tsouic languages 250–251
see also Austronesian languages; Formosan languages
- Tswana (Setswana) 1017
 Botswana 1017–1018
 noun classes 1018–1019
 South Africa 1017–1018
see also Sotho-Tswana languages
- Tswa-Ronga languages 1017
see also Bantu languages, Southern
- Tuareg 477
- Tuaregs, languages 152
- Tubar 1140
see also Uto-Aztecan languages
- Tubatulabal 1140
 classification 1139t
see also Uto-Aztecan languages
- Tucano
 adjectives 1099
 consonants 1094t
 morphemes 1096
 noun classifiers 1098
 speaker numbers 1092t
 syllable pattern 1095
- Tucanoan languages 1091–1103
 case markers 1096
 classification 1091
 classifiers 1097, 1098, 1098t, 1099
 nouns *see below*
 consonants 1092
 demonstrative adjectives 1099
 evidentiality 1100
 grammar 1096
 interrelationship 1092
 marriage aspects 1092
 iterative 1099–1100
 nasal spreading 1092, 1095–1096
 noun classifiers 1097
 animate 1098t
 inanimate 1098
 noun modifiers 1098
 adjectival verbs 1098
 adjectives 1099
 limiting adjectives 1098
 nouns 1097
 animate 1097
 classifiers *see above*
 inanimate 1097
 modifiers *see above*
 plurals 1097
 OVS 1096
 personal pronouns 1099
 progressive 1100
 sentence structure 1096
 suprasegmentals 1095
 accent 1096
 morphemes 1096
 nasal assimilation 1095
 nasalization 1095
 tone 1096
 syllable patterns 1095
 Tariana *vs.* 1051
 use of 1091
 verbs 1099
 auxiliary verbs 1100
 compound verb roots 1100
 evidentiality 1100
 future tense 1101
- vowels 1092
 word order 1096
see also Arawak languages; Tucanoan languages
- Tugbeni 517
- Tukanoan 750
see also Native American languages
- Tule 224
see also Chibchan languages
- Tulu, Malayalam *vs.* 682
- Tum 728–729
see also Cuoi languages
- Tumshuqese 541
see also Iranian languages
- Tundra *see* Yukaghir
- Tundra Nenets 761–762, 762–763
see also Nenets (Yurak)
- Tunebo
 borrowing from Spanish 651–652, 653
 Tunebo 651–652, 653
see also Chibchan languages
- Tungus 614
- Tungusic languages 1103–1105
 adjectives 1104
 Altaic hypothesis 653
 classification 250, 1103
 as endangered languages 1103
 genetic affiliation 1103
 morphology 1104, 1105t
 origin/development 1103
 phonology 1104
 consonants 1104, 1104t
 vowels 1104, 1104t
 SOV 1104
 structure 1104
 SVO 1104
 Turkic-Mongol relationship 30
 types 1103t
 use of 1103
 number of speakers 1103t
 verbs 1104
 writing systems 1103
see also Altaic languages; Evenki; Mongolia; Yakut
- Tunica 749
see also Muskogean languages
- Tunisia, Berber 152–153
- Tupari
 classification 1106t
 ideophones 1106–1107
 tone system 1106
see also Tupian languages
- Tupian languages 750
 adjectives 1106
 augmentative 1106
 case marking 1106
 classification 1105
 branches 1105–1106
 classifiers 1108
 core cases 1106
 diminutives 1106
 evidentiality 1108
 ideophones 1106–1107
 morphology 1106
 noun classification 1108
 nouns 1106
 phonetics 1106
 phonology 1106
 stress 1106
 tone system 1106
 positional demonstratives 1106
 postpositions 1106
 SOV 1107–1108
 syntax 1107–1108
 verbs 1106
 word classes 1106
 word order 1107–1108
see also Akuntsu; Arikém; Arúa; Awetí; Ayurú; Cariban languages; Guaraní; Macro-Jê languages; Native American languages
- Tupi-Guarani 1105–1106
 classification 1107t
 lexicon 1106

- Tupi-Guarani (*continued*)
 morphology 1106
 core cases 1106
see also Tupian languages
- Turi 736
see also Munda languages
- Turkey
 Arabic 42
 Aramaic 58
 Armenian 68
 Georgian 442
 Indo-Iranian languages 531
 Iranian languages 537
 Kurdish 538, 625
 official language 1112
 Turkish 1112
- Turkic languages 610, 1109–1112
 Altaic hypothesis 653
 classification 250, 1109
 contacts 1110
 development 1109
 written sources 1109
 features 1111
 loanwords 1112
 Mongol-Tungusic relationship 30
 morphology 1111
 Northwestern (Kipchak) branch 1109
 possessives 1112
 sound harmony phenomenon 1111
 Southwestern (Oghuz) branch 1109
 SOV 1111–1112
 syntax 1111–1112
 vowels 1111
 word accent 1111
 written varieties 1110
 contact effects 1111
 Karakhanid 1110
 Old Kirghiz 1110
 Old Uyghur 1110
 scripts 1110
 Volga Bulgar 1110
see also Altaic languages; Arabic;
 Azerbaijani; Bashkir; Chuvash; Iranian
 languages; Kazakh; Kirghiz; Mongolia;
 Nivkh; Slavic languages; Tajik Persian;
 Tatar; Tocharian; Turkish; Türkmen;
 Uralic languages; Uyghur; Uzbek; Yakut
- Turkish 120, 1112–1116
 auxiliary suffixes 1115
 future 1114–1115
 influence on other languages
 Abkhaz 2
 Azerbaijani 110–111
 Hindi 495–496
 New Iranian languages 538
 morphology 1114
 noun paradigm 1114
 as official language, Turkey 1112
 origin/development 1112
 phonology 1113
 consonants 1113, 1113*t*
 phonemes 1113
 rules 1114
 stress 1113
 vowels 1113, 1113*f*, 1114, 1114*t*
 possessives 1116
 postpositions 1113–1114
 pro-drop 1116
 progressive 1114–1115
 related languages, Azerbaijani 110–111
 SOV 1115–1116
 SVO 1115–1116
 syntax 1115
 use of 1112
 verb paradigm 1114
 vocabulary 1112–1113
see also Altaic languages; Arabic;
 Azerbaijani; Balkan linguistic area;
 Turkic languages; Türkmen
- Türkmen 1116–1119
 converbs 1119
 dialects 1119
 distinctive features 1117
 grammar 1118
 lexicon 1119
 location 1116
 origin/history 1117
 phonology 1117
 consonants 1118
 vowels 1117–1118
 related languages 1117
 Azerbaijani 110–111
 use of 1112, 1116
 written language 1117
 Arabic script 1117
 Cyrillic alphabet 1117
 Roman script 1117
see also Altaic languages;
 Azerbaijani; Persian, Modern;
 Russian; Turkic languages; Turkish;
 Uzbek; Yakut
- Turkmenistan
 Balochi 134
 Kazakh 588
 Uzbek 1145
- Tuscarora 542
see also Iroquoian languages
- Tutelo 749
see also Siouan languages
- Tutonish 75
- Tuvan 1200
- Tuyuca
 accent/tone 1096
 case markers 1097
 consonants 1094*t*
 demonstrative adjectives 1099
 evidentiality 1100
 morphemes 1096
 noun classifiers 1098
 noun modifiers 1099
 speaker numbers 1092*t*
 syllable pattern 1095
 verbs
 auxiliary verbs 1095
 evidentiality 1100
 vowels 1092
see also Tucanoan languages
- Twi (Akan)
 Akan 17
 Krio, influence on 620
- Tzeltalan 705–706
 speaker numbers 707*t*
see also Mayan languages
- Tzotzil 705–706
 long-range comparison 653
 speaker numbers 707*t*
see also Mayan languages
- Tz'utujil 705–706
 positionals 707
 speaker numbers 706*t*
see also Mayan languages
- ## U
- Ubykh
 phonemes 193
 vowels 193*t*
see also Caucasian languages
- Udi
 vowels 194, 194*t*
see also Caucasian languages
- Udmurt (Votyak)
 classification 1129–1130
 object marking 1132
 verbs 1131
 word order 1132
 word stress 1131
see also Permic (Permian) languages
- Uganda
 Luganda 657
 Luo 658
 Swahili 1026
- Ugaritic 932, 1121–1122
 alphabet 1121
 classification 250
see also Afroasiatic languages; Eblaite; Semitic
 languages
- !Ui-Taa languages 601–602
see also Khoesaa languages
- Ukaan–Apes
 classification 151
see also Benue-Congo languages
- Ukraine
 Hungarian 514
 Ukrainian 1122
- Ukrainian 1122–1123
 Belorussian *vs.* 147, 1122–1123
 classification 251–252, 974–975
 consonants 1122–1123
 Cyrillic alphabet 1122
 distinguishing features 1122
 nominal cases 1123
 Russian *vs.* 1122–1123
 use of 1122
 Ukraine 1122
 Yiddish, influence on 1205
see also Belorussian; Russian; Slavic languages
- Ulster, Irish, development of 454
- Ulwa 711
see also Misumalpan languages
- Umbrian language, Italic languages 555
- Umbuygamu *see* Morrobalama
- Ume Saami 911
see also Saami
- Unami
 classification 24
 origin/development 28
see also Algonquin languages
- UNESCO *see* Ad Hoc Expert Group on
 Endangered Languages
- United Kingdom (UK)
 Gujarati 468
 Urdu 1133
- Upernavik 1172
see also West Greenlandic
- Upper Chehalis 749
see also Salishan languages
- Upper German 445
- Ural-Altaic languages 30
- Uralic languages 1129–1133
 aspect 1131–1132
 case suffixes 1131
 classification 249, 1129–1130
 definiteness 1131
 gender 1131
 morphology 1131
 negation 1131
 Nostratic theory 249, 653–654, 786
 objects 1132
 phonology 1130
 consonant gradation 1130
 consonantism 1130
 diphthongs 1130
 vocalism 1130
 vowel harmony 1130
 word stress 1131
 plural markers 1131
 postpositions 1131–1132
 SOV 1132
 subordinate sentences 1132
 syntax 1131
 use of
 geographical distribution 1129
 Russia 1129–1130
 verbs 1131
 main verb phrases 1132
 vowel harmony 1130
 word order 1132
see also Altaic languages; Estonian; Finnish
 (Suomi); Hungarian; Language
 endangerment; Mongolia; Nenets (Yurak);
 Saami; Turkic languages; Yukaghir
- Urban centers, and language endangerment 325,
 326
- Urban sign languages 954
- Urdu 1133–1139
 classification 251–252, 522
 codification/standardization 1136
 conflict with Hindi 1134
 dictionaries 1136
 grammar 1135

- Hindustani, divergence from 1135–1136
 influence on other languages
 Burushaski 179
 Hindustani 497
 Kashmiri 582–583
 Malayalam 680–681
 Punjabi 889
 Telugu 1058
 lexical borrowing 1135
 literature 1134, 1137
 Islamic traditions 1137
 Persian influences 1137
 as national language, Pakistan 1133
 number of speakers 522–523
 as official language, Pakistan 885–886
 origin/development 1133
 as literary language 1133
 Persian influences 1134
 script 1134
 written records 1133
 popularization 1136
 use of 1133
 Bangladesh 522–523, 1133
 India 522–523, 1133
 Pakistan 522–523, 1133
 vocabulary 1135
 writing systems 524
see also Arabic; Dardic languages; Hindi;
 Hindustani; Indo-Aryan languages; Pashto
- Uribe, José Vincente, Emberá studies 226
- Urmia 112–113
see also Azerbaijani
- Uru 752
see also Native American languages
- USA 1123–1129
 African-American English 1125
 American Creoles 1127
 American English 1123
 American Sign Language 1127
 bilingualism debate 1125
 Creoles 1127
 Cupeño 270
 Dutch 307
 English 1123
 Fijian 412
 Finnish 413
 Gullah 470
 Hawaiian Creole English 1127
 Inupiaq 535
 Italian 545
 Louisiana Creole French 656, 1127
 Michif 709
 minority immigrant languages 1127
 Native American languages 1127
 Omaha-Ponca 802
 Spanish 1020, 1123, 1126
 Tohono O'odham 1074
 used at home 1124*f*
see also African-American Vernacular
 English (AAVE); Algonquin Languages;
 Algonquin languages; Caddoan
 languages; Central Siberian Yupik;
 Creek; Creoles; Cupeño; English;
 Eskimo-Aleut languages; Hoka
 languages; Hopi; Inupiaq; Iroquoian
 languages; Keres; Lakota; Michif;
 Muskogean languages; Na-Dene
 languages; Nahuatl; Navajo;
 Omaha-Ponca; Oneida; Pidgins;
 Polysynthetic languages; Pomoan
 languages; Ritwan languages; Salishan
 languages; Sign language; Siouan
 languages; Tohono O'odham; Uto-Aztecan
 languages; Wakashan languages
- Uspanteko 705–706
 speaker numbers 706*t*
see also Mayan languages
- Uto-Aztecan languages 748
 classification 1139
 grammar 1140–1141
 internal relationships 1140
 phonology 1141
 records/studies 1140
 Tanoan language link 1140
- texts 1140–1141
 use of 1139
 workers in 1140
see also Aztec; Cupeño; Hopi; Nahuatl;
 Native American languages; Tohono
 O'odham
- Utoro 613
see also Kordofanian languages
- Uyghur 1142–1145
 contacts 1143
 dialects 1144
 distinctive features 1143
 evidentiality 1144
 grammar 1144
 lexicon 1144
 location 1142
 origin/history 1142
 phonology 1143
 vowels 1143–1144
 possessives 1144
 related languages 1143
 Uzbek 1146
 speakers 1142
 use of 1142
 Xinjiang 1142
 written language 1143
 Arabic script 1143
 Cyrillic script 1143
 Roman script 1143
see also Altaic languages; Kazakh; Kirghiz;
 Turkic languages; Uzbek
- Uyghur, Old 1110
see also Turkic languages
- Uzbek 1145–1148
 contacts 1146
 converbs 1147
 dialects 1147
 evidentiality 1147
 grammar 1147
 lexicon 1147
 location 1145
 Russian bilingualism 1145
 origin/history 1145
 phonology 1146
 consonants 1147
 sound harmony 1147
 suffixes 1147
 Tajik Persian *vs.* 1041
 vowels 1146
 related languages 1146
 Kazakh 589
 Uyghur 1146
 use of 1145
 vowel harmony 1147
 written language 1146
 Arabic script 1146
 Cyrillic script 1146
 Roman script 1146
see also Altaic languages; Kazakh; Tajik
 Persian; Turkic languages; Türkmen;
 Uyghur
- Uzbekistan
 Kazakh 588
 Kirghiz 610
 Russian 1145
 Tajik 1145
 Tajik Persian 1041
 Uzbek 1145
- V**
- Vai, Krio, influences on 620
- Valency
 Evenki 407
 Romani 899
 Tamambo 1046
- Vanuatu
 Bislama 161
 English 161
 French 161
 Vurës 1154
- Van Wyk Louw, N P, Afrikaans 9–10
- Variable order, postbases 202, 203
- Variation theory, African-American Vernacular
 English development 337
- Varma, A A RajaRaja, Malayalam 681
- Vatican City, Italian 545
- Vatteluttu 681
- Vedas
 nouns 533
 word order 534
see also Hinduism
- Venda 1017
see also Bantu languages, Southern
- Venetian *see* Italian
- Venezuela
 Andean languages 40
 Arawak languages 59
 Cariban languages 40
 Chibchan languages 40
 Guajiro 59
- Veps 1129–1130
see also Finnic languages
- Verb(s)
 agreement, sign language 943*f*
 derivation, postbases 202
 directional 1014
 inflections
 in agglutinating languages 417
 suffixes 221
 introflecting languages, stems 50
 modification 951
 in polysynthetic languages 203
 Tanoan languages 1049–1050
see also specific languages
- Verbal predicates, Thai 1059–1060
- Verb-final languages 288
- Verb-initial languages 288
- Verb-medial languages 288
- Verb-medial sentence type, Karen languages 581
- Verificationality *see* Evidentiality
- Vernacular Hindustani 499
- Verner's Law 449*t*
- Versatility, Southeast Asian languages 1012
- Viet-Muong languages 724
 use of 728–729
see also Mon-Khmer languages
- Vietnam
 Bahnaric languages 725–726
 Katuic languages 726–727
 Khmuic languages 727
 Mang languages 729
 Mon-Khmer languages 725
 Tai-Kadai languages 105
 Tai languages 1039
 Viet-Muong languages 728–729
- Vietnamese 1149–1154
 Chinese, influence from 248, 728–729, 1149
 classification 250
 consonants 1150, 1150*t*
 coverbs 1014
 dictionaries 1152
 grammars (books) 1152
 grammaticalization 1012
 historical origins 1149
 orthography 1149–1150
 as isolating language 291
 phonology 1150
 phrases and sentences 1152
 regional varieties 1150
 Central (Huế) 1150
 Northern (Hanoi) 1150
 Southern (Hô Chí Minh City) 1150
 script 1149
 sources 1152
 syllable rhymes 1151, 1151*t*
 tones 1150, 1150*t*
 use of 728–729
 word category and construction 1151
 compounds 1151–1152
 word order 1014*t*
 word structure 1150–1151
 workers in 1149–1150
see also Mon-Khmer languages; Viet-Muong
 languages
- Viking *Bund* 392–393
- Village-based sign languages 954

- Vinaya Pitaka*, Pāli canonical texts 831–832
Vocabulario de la lengua Bicol 158
Vocabulario de la Lengua Bisaya 915
 Vocabulary inspection, Native American languages 747
 Vocalic melody, in introflecting language 51
 Voiced stops, Quileute 211
 Volapük 76
 Volga Bulgar 1110
 see also Turkic languages
 Volga-Kama *Sprachbund* 392–393
 Voltaic languages *see* Gur (Voltaic) languages
 Von der Gabeltzen, H.C., Austronesian languages 98
 Von le Coq, Albert, Tocharian 1068–1069
 Votayk *see* Udmurt (Votayk)
 Votic
 classification 1129–1130
 postposition 290
 see also Finnic languages
 Vowel(s)
 harmony *see* Vowel harmony
 length, European linguistic area 397*f*
 mutation, in agglutinating languages 418
 raising 122
 rounded 396*f*
 see also specific languages
 Vowel harmony
 in agglutinating languages 417–418
 Akan 19
 Azerbaijani 111
 Bashkir 143–144
 Chukotko-Kamchatkan languages 240
 Dogon 294
 Finnic languages 1130
 Finnish (Suomi) 414, 417–418
 Ga-Dangme 632
 Gbe languages 632
 Gur (Voltaic) languages 472
 Hungarian 514–515, 1130
 Kashmiri 583
 Khanty 1130
 Kinyarwanda 605
 Kunama 773–774
 Kwa languages 632
 Luo 658–659
 Macro-Jê languages 667
 Madurese 673
 Mansi 1130
 Mari languages 1130
 Mordvin languages 1130
 Nganasan (Tavgy) 1130
 Nilo-Saharan languages 773–774
 Nilotic languages 773–774
 Nubian 773–774
 Songai languages 773–774
 Tano languages 632
 Telugu 1055
 Temein 773–774
 Tibetan 1062
 Togo Mountain languages 632
 Turkish 1114
 Uralic languages 1130
 Uzbek 1147
 Yoruba 1208
 VSO
 Akkadian 21
 Arabic 47
 Berber languages 154
 Chadic languages 207
 Chinantecan languages 211
 Egyptian 39
 Finnish (Suomi) 413
 Ge'ez 442
 Hawaiian Creole English (HCE) 480–481
 Maori 700
 Mixe-Zoquean languages 715
 Munda languages 737
 Niuean 776
 Nuuchahnulth (Nootka) 790
 Oto-Mangean languages 822
 Samar-Leyte 913
 Scots Gaelic 927
 Tai languages 1039
 Tohono O'odham 1075
 Totonacan languages 1084
 Wakashan languages 1160
 Welsh 1171
 Yanan languages 508
 Zapotecan languages 1214
 'Vulgar Latin,' 641
 Vurës 1154
 classifiers 1154
 consonants 1154
 as endangered language 1154
 as nominative-accusative language 1154
 phonetics 1154
 possession marking 1154
 use of, Vanuatu 1154
 verb serialization 1154–1154
 vowels 1154
 word order 1154
 see also Austronesian languages
- W**
- Wa 1155–1157
 Bible translation 1155
 classification 250
 influence from other languages 1156, 1156*t*
 as isolating language 1156
 literacy 1156
 orthography 1155, 1155*t*
 personal pronouns 1156, 1156*t*
 prefixes 1156, 1156*t*
 syllable-initial consonants 1155, 1155*t*
 syntax 1156
 use of 1155
 vowel registers 1156*t*
 vowels 1155–1156, 1156*t*
 vowel registers 1155–1156
 see also Austroasiatic languages; Mon-Khmer languages; Waic languages
 Wackernagel's Law, Latin syntax 643
 Wahl, Edgar de, artificial languages 77
 Waic languages 727
 classification 728
 nomenclature 728
 use of 728
 see also Palaung-Wa languages
 Waimaja/Bará
 accent/tone 1096
 consonants 1094
 speaker numbers 1092*t*
 verbs 1099–1100
 see also Tucanoan languages
 Waimiri-Atroari
 geographical distribution 185*f*
 phonology 183–184
 see also Cariban languages
 Waiwai
 geographical distribution 185*f*
 phonology 183–184
 see also Cariban languages
 Waja languages 3
 see also Adamawa-Ubangi languages
 Wakashan languages 749
 Chimakuan family 750
 classification 1157
 Northern group 1157
 Southern group 1157
 classifiers 1159–1160, 1160*t*
 compounding 1160
 consonants 1158
 diminutives 1159
 glottalization 1158, 1159*t*
 Kwakiutlan branch 749–750
 lenition 1158, 1159*t*
 morphology 1159
 nominal phrase 1160
 Nootkan branch 749–750
 Northern *vs.* Southern groups 1158
 person-number inflections 1160
 phonology 1158
 possession 1160
 Proto-Wakashan 1158, 1158*t*
 reduplication 1159, 1159*t*
 stress assignment 1158
 suffixes 1159, 1159*t*
 syllable structure 1158
 syntax 1160
 tense markers 1160
 use of 1157
 speaker numbers 1157–1158
 vowels 1158
 vowel epenthesis 1158–1159
 VSO 1160
 word order 1160
 see also Areal linguistics; Native American languages; Nuuchahnulth (Nootka)
 Wambaya 1161–1165
 adjectives 1162–1163
 case marking 1164
 cases 1163, 1163*t*
 classification 250
 clauses 1164
 concord 1162
 definition 1161
 earliest records 1161–1162
 ergative 1163
 genders 1163
 morphology 1162
 nouns 1162–1163
 numbers 1163
 phonemes 1162, 1162*t*
 phonology 1162
 possessives 1163
 prefixes 1162
 stops 1162
 switch reference 1164
 syntax 1164
 use of 1161
 verb-headed clauses 1163–1164
 verbs 1162–1163
 word order 1164
 word structure 1162
 see also Australian languages
 Wanano
 adjectival verbs 1099
 consonants 1095
 evidentiality 1100
 speaker numbers 1092*t*
 verbs, evidentiality 1100
 see also Tucanoan languages
 Wancho
 classification 968–969
 see also Konyak languages
 Waray-Waray
 diminutives 915–916
 future 915–916
 progressive 915*t*
 see also Samar-Leyte
 Warekena (Guarequena) 60
 see also Arawak languages
 Warlpiri 1165–1169
 auxiliary complexes 1167, 1167*t*
 cases 1166–1167
 clauses 1166, 1166*t*
 consonants 1165, 1165*t*
 dialects 1165
 dictionaries 1165
 as ergative languages 88
 Hale, Ken 1165
 imperfective 1166*t*
 iterative 1166
 meaning/context 1167
 counting system 1167*t*, 1168
 kin terminology 1168
 spatial orientation 1167–1168, 1167*t*
 mora counting rule 1166
 morphology 1165
 nouns 1166, 1167*t*
 phonology 1165
 respect 1168
 spatial cases 1166–1167, 1167*t*
 suffixes 1166
 use of 1165
 speaker numbers 1165
 verbal clauses 1166
 verbs 1166, 1166*t*
 vowels 1165
 word structure 1165

- see also* Arrernte; Australian languages;
 Kaytetye; Pama-Nyungan languages;
 Pitjantjatjara; Sign Language
 Warnang 613
see also Kordofanian languages
 Warriangka 570
 Warupu (Barupu) 974
see also Skou languages
 Washo 750–751
see also Hoka languages
 Washu languages 504
 classification 505
see also Hoka languages
 Wasteko 705–706
see also Mayan languages
 Waunana *see* Choco languages
 Waunméu
 classification 224
 use of 224
see also Choco languages
 Waurá 60
see also Arawak languages
 Wayana
 geographical distribution 185*f*
 reduplication 184
 vowels 183–184
see also Cariban languages
 Waygali 787
see also Nuristani languages
 Waziri metaphony, Pashto 845–846
 Weak grade, in agglutinating languages 418, 418*t*
 Wedebo 624
see also Grebo languages
 Weinreich, Max, Yiddish development 1205
 Welsh 1169–1172
 alphabet 1169
 classification 200, 251–252
 demography 1169
 dialects 1170*t*
 mutation 1170
 periods 1169
 phonemes 1169
 consonant mutation 1170, 1170*t*
 consonants 1169–1170, 1170*t*
 diphthongs 1169–1170, 1170*t*
 vowels 1169–1170, 1170*t*
 stylistic variation 1171
 syntax 1171
 use of 1169
 decline of 1169
 geographic variation 1169
 Patagonia 1169
 revival 1169
 speaker numbers 1169
 vocabulary 1170
 Celtic roots 1170
 dialects 1170–1171
 English loanwords 1170
 VSO 1171
 word order 1171
see also Breton; Brythonic Celtic; Celtic;
 Cornish; Pictish; Scots Gaelic
 Welsh Language Board 1169
 West Bird's Head (WBH) languages 1176
see also West Papuan languages
 West Bomberai languages
 classification 1087
see also Trans New Guinea languages
 Westermann, Diedrich Hermann
 Adamawa-Ubangi languages 2
 Gur language studies 472
 Kwa language classification 630
 Mande language classification 696–697
 Niger-Congo languages 768
 Western Bwe 581
see also Karen languages
 Western Gaelic 454
 Western Kru 624
 Ivory Coast 624
 Liberia 624
see also Kru languages
 Western Sahara, Spanish 1020
 Western Songai, SVO 991
 Western Yiddish 1206
 West Germanic *see* Germanic languages
 West Greenlandic 1172–1176
 autolexical theory 1173–1175
 Bible translations 1175
 classification 251, 1172
 consonants 1172–1173, 1173*t*
 Danish, influences from 1175
 discourse 1175
 ergativity 1173–1175
 lexicon 1175
 morphology 1173
 nouns 1173
 phonetics/phonology 1172
 roots 1173
 semantics 1175
 sociolinguistics 1175
 stops 1172–1173
 syntax 1173
 transitivity 1173–1175
 use of 1175
 geographical distribution 1172, 1174*f*
 verbs 1173
 vowels 1172–1173
see also Eskimo-Aleut languages; Greenlandic
 (Kalaallit); Inupiaq
 West New Britain languages 841
see also Anem; Ata
 West Papuan languages 1176–1179
 classification 253, 1176
 contact with other languages 1177
 nominal complex 1177
 genders 1177
 noun phrase 1177
 numbers 1177
 pronominal system 1177
 SOV 1176
 SVO 1177–1178
 tone 1177
 use of 1176
 geographical distribution 1177*f*
 verbal complex 1176
 negative adverbs 1176–1177
 Tense-Mood-Aspect 1176
 word order 1176
see also Abun; Austronesian languages; Papuan
 languages; Trans New Guinea languages
 West Saxon, Old English dialect 356
 West Semitic languages *see* Semitic languages,
 West
 West Siberian Tatar 1052
 Westtocharisch *see* Tocharian
 Wexler, Paul, Yiddish development 1205
 White Tai (Tai Dón) 1039
see also Tai languages
 Whorf, Benjamin Lee
 Hopi 511
 Uto-Aztecan languages 1140
 Wichita 749
see also Caddoan languages
 Wilkins, John, artificial languages 76
 Will/have future tense, Balkan linguistic area 128,
 129*t*
 Williamson, Kay, Mande language classification
 696–697
 Wintun 750
see also Penutian languages
 Wissel Lakes languages 1087
see also Trans New Guinea languages
 Wiyot
 classification 25
 long-range comparisons 651
see also Algonquin languages
 Wobe 624
see also Guere languages
 Wolaitta 1179–1184
 adjectives 1181
 adverbs 1181
 clauses 1183
 complex clause 1183
 simple declarative clause 1183
 consonants 1179, 1180*t*
 diminutives 1180
 family tree 1179*f*
 future 1182
 imperfective 1182
 nouns 1180, 1180*t*
 case 1180, 1181*t*
 definiteness 1180
 derivation 1180, 1181*t*
 gender 1180, 1181*t*
 plurals 1180, 1180*t*, 1181*t*
 phonology 1179
 possessives 1181–1182
 pronouns 1181, 1182*t*
 gender 1181–1182, 1182*t*
 SOV 1183
 syllable structure 1180
 tone-accent 1180
 use of 1179
 verbs 1182
 aspect 1182, 1182*t*
 imperative mood 1183, 1183*t*
 interrogatives 1182, 1183*t*
 modality 1182
 negation 1182, 1182*t*
 subject agreement 1182, 1182*t*
 vowels 1179, 1180*t*
see also Afroasiatic languages; Ethiopian
 linguistic area (ELA); Omotic languages
 Wolof 770
 advanced tongue root (ATR) feature
 1184–1185
 classification 253
 classifiers 1185–1186
 consonants 1184–1185
 genetic affiliation 1184
 Sapir 1184
 ideophones 1184–1185
 influence on other languages, Krio 620
 influences from other languages, French
 1186–1186
 morphology 1185
 noun class 1185–1186
 phonetics/phonology 1184
 stops 1184–1185
 syntax 1185
 urban type 1186
 use of 1184
 verbs 1185–1186
 vowels 1185*f*
see also Atlantic Congo languages
 Woordenboek der Nederlandse taal (WNT)
 308
 Word(s)
 accent 1111
 borrowing 980
 order *see* Word order
 Word formation, diminutives *see* Diminutives
 Word order
 Afrikaans 9
 Balkans *see* Balkan linguistic area
 language diffusion 248
 morphological types 733–734
 Tanoan languages 1049–1050
see also SOV; SVO; VSO
 Word stress
 Kaytetye 587
 Polish, phonology 875
 Romanian 901–902
 Saami 912
 Slovak 978
 Uralic languages 1131
 Yiddish 1204
 World Englishes 363–371
 'concentric circles' model 364, 364*f*
 creativity 368
 definition 363
 literature 368
 nativization 365
 speech communities 364
see also Bilingualism; English; English, Modern
 World Esperanto Conference 375
 Writing/written language
 Bashkir 143
 Karen languages 581
 sign language *see* Sign language
 Urdu 524
 Written Mongol 723

Wu languages 219
 classification 969
 speaker numbers 214*t*
see also Chinese
 Wulfilā *see* Gothic
 Wu-ming (Northern Zhuang) 1039
see also Tai languages
 Wurm, S A, Gamilaraay study 438
 Wyld, Henry C, Later Modern English
 definition 343

X

Xhosa 1017
 aspect morphemes 1191
 augmentative 1188
 classification 253
 clicks 1018
 diminutives 1188*t*
 Fanagalo, influences on 411
 ideophones 1197
 intransitive 1197
 transitive 1197
 imperfective 1192
 mood inflexion 1193
 consecutive mood 1196
 imperative mood 1196
 indicative mood 1193
 infinitive mood 1197
 participle (situative) mood 1193
 relative mood 1193
 subjunctive mood 1194
 temporal mood 1196
 negative inflexion 1192
 nouns 1187
 agreement morphology 1188
 classes 1187
 compound 1188
 derived nouns 1188
 noun classes 1193
 suffixes 1187
 as official language, South Africa 1018
 perfect 1192
 possessives 1188
 progressive 1191
 speaker numbers 1018
 verbal inflection 1191
 subject/object agreement prefixes 1191
 verbs 1189
 causative suffix 1190
 compound past tense 1192
 derivation 1189
 detransitivising affixes 1190
 future tense 1192
 perfect past tense 1192
 present tense 1191
 recent compound past tense 1192
 remote compound past tense 1192
 remote past tense 1192
 tense inflexion 1191
 transitivity 1189
 unaccusative suffixes 1190
 Zulu *vs.* 1215
see also Bantu languages; Bantu
 languages, Southern; Fanagalo;
 Nguni languages; Niger-Congo
 languages; Zulu
 Xiang languages
 classification 969
 speaker numbers 214*t*
see also Chinese
 Xinca 748
see also Native American languages
 Xinjiang 1142
 Iranian languages 537
 Xipáya
 classification 1106*t*
 ideophones 1106–1107
 tone system 1106
see also Tupian languages
 Xitsonga
 South Africa 1187
see also Tonga

Xoklém 668
see also Jê languages
 Xoy 112–113
see also Azerbaijani
 !Xung languages 601–602
see also Khoesaa languages

Y

Yaghan (Yámana) 41
 Mapudungan languages 701
see also Andean languages
 Yaghnobi 541
see also Iranian languages
 Yahgan *see* Andean languages
 Yahudic *see* Judeo-Arabic
 Yakan 1004
see also South Philippine languages
 Yakut 1199–1202
 contacts 1200
 Evenki 1200
 Yeniseian 1200
 dialects 1201
 grammar 1200
 lexicon 1201
 location 1199
 Novgorodov, S A 1200
 origin/history 1199
 phonology 1200
 vowels 1200
 possessives 1200
 related languages 1200
 Khakas 1200
 Tuvan 1200
 sound harmony 1200
 use of 1199
 written language 1200
 Cyrillic alphabet 1200
see also Altaic languages; Tungusic languages;
 Turkic languages; Türkmen
 Yalarnnga
 as ergative languages 88
see also Australian languages
 Yalunka 620
 Yámana *see* Yaghan (Yámana)
 Yanan languages 750–751
 classification 505
 person markers 507
 VSO 508
see also Hokan languages
 Yanito 1202–1203
 classification 249*t*
 code switching 1202
 diminutives 1202
 influence from other languages
 Arabic 1202
 English 1202
 Hebrew 1202
 Italian 1202
 Spanish 1202
 perfect 1202
 use of 1202
 Gibraltar 1202
see also Creoles; Pidgins; Spanish
 Yankunytjatjara *see* Pitjantjatjara
 Yareba languages 1087
see also Trans New Guinea languages
 Yatè
 classification 665, 666, 666*t*
 geographical distribution 666–667
 inflectional morphology 667
 vowels 667
see also Macro-Jê languages
 Yawa
 classification 1176
 word order 1176
see also West Papuan languages
 Yawalapiti 60
see also Arawak languages
 Yawarana 185*f*
see also Cariban languages
 Yaz'va-Komi 1129–1130
see also Permic (Permian) languages
 Yega 613
see also Kadugli languages
 Yeme'an 506
see also Hokan languages
 Yeniseian 1200
 Yenisey-Samoyed *see* Enets (Yenisey-Samoyed)
 Yerwa *see* Kanuri
 Yeshvish *see* Judeo-English
 Yiddish 565, 567, 1203–1206
 adverbs 1204
 consonants 1203–1204
 development 567
 dialects 1206
 Eastern 1206
 Standard 1206
 Western 1206
 diminutives 1204
 diphthongs 1203–1204
 earliest documents 567
 future 1204
 history 1205
 population movement 1205–1206
 influence from other languages 1205
 Aramaic 567, 1205
 German 444, 1205
 Hebrew 567, 1205
 lexicon 1205
 morphology 1204
 noun genders 1204
 obstruents 1204
 orthography 1203
 Hebrew alphabet 1203
 oldest text 1206
 phonology 1203
 plurals 1204
 pronouns
 reflexive 1205
 subject 1205
 scholarship 565
 Standard 1206
 syntax 1204
 use of 566, 567, 1203
 speaker numbers 567
 verbs 1204
 vowels 1203–1204
 word order 1205
 as word-second language 1204
 word stress 1204
see also Germanic languages; Hebrew, Israeli;
 Jewish languages; Semitic languages
 Yidiny 88–89
 morphology/syntax 90
see also Australian languages
 Yinglish *see* Judeo-English
 Yokutsan 750
see also Penutian languages
 Yoruba 1207–1210
 assimilated low tone 1208
 Bible translation 1207
 consonants 1208
 dialects 1209
 earliest written records 1207
 example 1208–1209
 gender system 1208
 genetic relationships 1207
 grammars (books) 1207
 high tone restrictions 1208
 history 1207
 influence on other languages
 Hausa 477
 Krio 620
 Kwa languages *vs.* 631
 morphology 1208
 noun classes 1208
 number 1208
 orthography 1207
 Roman alphabet 1207
 past/present actions 1208
 phonetics/phonology 1208
 possessive noun–noun constructions 1208
 pronoun tone 1208
 syntax 1208
 use of 1207
 Benin 1207

- Zulu (*continued*)
 geographical distribution 1207f
 Nigeria 1207
 teaching of 1209
 Togo 1207
 verbal constructions 1208
 vowels 1208
 co-occurrence restrictions 1208
 elision 1208
 vowel harmony 1208
 workers in 1207
see also Defoid
- Yoruboid languages 151
see also Benue-Congo languages
- Young, Thomas, Indo-European languages 528
- Yucatecan 705–706
 noun classifiers 708
 speaker numbers 707t
see also Mayan languages
- Yuchi 749
see also Muskogean languages
- Yue languages 218
 classification 969
 speaker numbers 214t
see also Chinese
- Yugoslavia
 Slovak 977
 Turkish 1112
- Yukaghir 1210–1212
 case features 1210–1211
 classification 249
 consonants 1210–1211
 converbs 1211–1211
 ‘focus marking,’ 1210–1211
 history 1210
 nasalization 1210–1211
 SOV 1211–1211
 syntax 1211–1211
 use of 1210
see also Uralic languages
- Yukian 749
see also Muskogean languages
- Yukpa
 geographical distribution 185f
 stress 183–184
see also Cariban languages
- Yukuben 253
see also Benue-Congo languages
- Yuman languages 504, 750–751
 classification 506
 person markers 507
see also Hokan languages
- Yungur languages 3
see also Adamawa-Ubangi languages
- Yupik 373
 history 371
 Russian, influences from 373
 syntax 373
 use of 373
see also Eskimo-Aleut
- Yuracaré 41
see also Andean languages
- Yurak *see* Nenets (Yurak)
- Yurats 1129–1130
see also Samoyed languages
- Yurok
 classification 25
 long-range comparisons 651
see also Algonquin languages
- Yurumagui *see* Barbacoan languages
- Yuruti
 accent/tone 1096
 consonants 1094t
 morphemes 1096
 noun modifiers 1098
 speaker numbers 1092t
 verbs 1101
see also Tucanoan languages
- Yuwaalaraay 439
- ## Z
- Zaborski, A, Ethiopian linguistic area (ELA) 379
- Zaire, languages, Adamawa-Ubangi languages 771
- Zaiwa 968–969
see also Lolo-Burmese languages
- Zakataly 112
see also Azerbaijani
- Zambia
 Fanagalo 411
 Lozi 1017–1018
 Nyanja 791
 Shona 938
- Zamenhof, Ludovic Lazar 375
 Esperanto 76–77
Linguo internacia 375
- Zande languages 3
see also Adamawa-Ubangi languages
- Zapotec 751
 classification 1213
 speaker numbers 1213
see also Zapotecan languages
- Zapotecan languages 751
 agreement 1214
 classification 1213
 consonants 1213
 documentation 1213
 morphology 1213
 noun classes 823
 phonology 1213
 ‘pied-piping with inversion,’ 1214
 position hierarchy 1214
 prefixes 1214
 pronouns 1214
 speaker numbers 1213
 syllable onsets 821
 syntax 1213
 as tonal language 1213
 verbs 1213
- VSO 1214
 word order 1214
see also Chatino
- Zarphatic *see* Judeo-French
- Zezuru 1017
see also Shona languages
- Zhejiang 969
see also Hui languages
- Zhongyuan
 classification 214
 speaker numbers 214t
see also Mandarin
- Zhuang-Dong *see* Tai-Kadai (Zhuang-Dong)
- Zimbabwe
 Fanagalo 411
 Nyanja 791
 Shona 938
 Shona languages 1017
 Southern Bantu languages 1017
 Tshwa 1018
 Venda 1017
 Zimbabwean Ndebele 1018
- Zimbabwean Ndebele 1018
 official language 1018
- Zimbabwe 1018
see also Nguni languages
- Zois, Sigismund, Slovene 981
- Zoque 713
see also Mixe-Zoquean languages
- Zoroastrianism
 Avestan 107
 Indo-Iranian languages 531–532
 Pahlavi (Middle Persian) 538, 827
- Zorque de Rayón *see* Mixe-Zoquean languages
- Zulu 1017
 classification 137
 clicks 1215–1216
 comparative linguistics 1216
 dialects 1216
 Fanagalo, influence on 412
 history 1216
 morphology 1215
 Ndebele *vs.* 1215
 nouns 1215
 noun classes 1017
 as official language, South Africa 1018
 phonology 1215
 sociolinguistics 1216
 SVO 1215–1216
 Swati *vs.* 1215
 syntax 1215
 use of 1215
 speaker numbers 1018
 word order 1215–1216
 Xhosa *vs.* 1215
see also Afrikaans; Bantu languages; Bantu languages, Southern; Fanagalo; Nguni languages; Xhosa
- Zuni 750
see also Penutian languages